GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 19.461 Seconds
(without alignments)
412.880 Million cell updates/sec

Title: US-09-834-309-4
Perfect score: 751
Sequence: 1 GISCGSPPPTINGRISYYST ... ANNMWGPTPLPTCVSVFPLE 184

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	219	29.2	1466	6	5256642 6	Patent No. 5256642
2	219	29.2	1466	6	5472939 6	Patent No. 5472939
3	219	29.2	1537	6	5256642-5	Patent No. 5256642
4	219	29.2	1537	6	5472939-5	Patent No. 5472939
5	219	29.2	1847	6	5256642 10	Patent No. 5256642
6	219	29.2	1847	6	5472939 10	Patent No. 5472939
7	219	29.2	2039	6	5256642 2	Patent No. 5256642
8	219	29.2	2039	6	5472939 2	Patent No. 5472939
9	214	28.5	263	1	US-07-906-983-2	Sequence 2, Appli
10	198	26.4	133	2	US-08-356-361-31	Sequence 31, Appl
11	198	26.4	133	2	US-08-769-967A-31	Sequence 31, Appl
12	198	26.4	254	2	US-08-356-361-29	Sequence 29, Appl
13	198	26.4	254	2	US-08-769-967A-29	Sequence 29, Appl
14	195	26.0	254	2	US-08-356-361-30	Sequence 30, Appl
15	195	26.0	254	2	US-08-769-967A-30	Sequence 30, Appl
16	190	25.3	463	4	US-08-981-234B-2	Sequence 2, Appli
17	160	21.4	484	4	US-08-139-195-2	Sequence 2, Appli
18	159	21.2	169	1	US-08-310-416A-18	Sequence 18, Appl
19	159	21.2	169	2	US-08-888-171-18	Sequence 18, Appl
20	159	21.2	254	1	US-08-310-416A-13	Sequence 13, Appl
21	159	21.2	254	2	US-08-888-171-13	Sequence 13, Appl
22	159	21.2	254	2	US-08-435-149-1	Sequence 1, Appli
23	159	21.2	293	1	US-08-310-416A-16	Sequence 16, Appl
24	159	21.2	293	2	US-08-888-171-16	Sequence 16, Appl
25	159	21.2	470	2	US-08-528-057-42	Sequence 42, Appl
26	159	21.2	473	2	US-08-528-057-44	Sequence 44, Appl
27	159	21.2	477	2	US-08-528-057-2	Sequence 2, Appli
28	159	21.2	484	6	5514787 2	Patent No. 5514787
29	159	21.2	577	2	US-08-435-149-3	Sequence 3, Appli
30	159	21.2	611	4	US-09-475-660A-32	Sequence 32, Appl
31	158.5	21.1	323	2	US-08-435-149-2	Sequence 2, Appli
32	158.5	21.1	324	1	US-08-310-416A-14	Sequence 14, Appl
33	158.5	21.1	324	2	US-08-888-171-14	Sequence 14, Appl
34	155	20.6	324	2	US-08-528-057-46	Sequence 46, Appl
35	153.5	20.4	197	2	US-08-356-361-27	Sequence 27, Appl
36	153.5	20.4	197	2	US-08-769-967A-27	Sequence 27, Appl
37	147	19.6	196	3	US-08-824-692-32	Sequence 32, Appl
38	147	19.6	229	3	US-08-824-692-31	Sequence 31, Appl
39	147	19.6	290	3	US-08-824-692-29	Sequence 29, Appl
40	144	19.2	126	6	5514582 43	Patent No. 5514582
41	144	19.2	1019	1	US-08-296-614A-4	Sequence 4, Appli
42	144	19.2	1019	2	US-08-536-405-4	Sequence 4, Appli
43	144	19.2	1019	2	US-08-687-620-4	Sequence 4, Appli
44	144	19.2	1083	1	US-08-296-614A-2	Sequence 2, Appli
45	144	19.2	1083	2	US-08-536-405-2	Sequence 2, Appli
46	144	19.2	1083	2	US-08-687-620-2	Sequence 2, Appli
47	140	18.6	274	2	US-08-177-109A-58	Sequence 58, Appl
48	140	18.6	274	2	US-08-687-706-58	Sequence 58, Appl
49	137.5	18.3	86	6	5514582-47	Patent No. 5514582
50	133	17.7	124	6	5514582-48	Patent No. 5514582
51	133	17.7	265	2	US-08-177-109A-57	Sequence 57, Appl
52	133	17.7	265	2	US-08-687-706-57	Sequence 57, Appl
53	133	17.7	764	2	US-08-177-109A-2	Sequence 2, Appli
54	133	17.7	764	2	US-08-687-706-2	Sequence 2, Appli
55	129	17.2	177	3	US-08-824-692-36	Sequence 36, Appl
56	127	16.9	248	2	US-08-640-977-2	Sequence 2, Appli
57	127	16.9	266	2	US-08-640-977-4	Sequence 4, Appli
58	127	16.9	326	2	US-08-640-977-1	Sequence 1, Appli
59	126.5	16.8	128	6	5514582-44	Patent No. 5514582
60	126	16.8	129	6	5514582-49	Patent No. 5514582
61	125.5	16.7	76	2	US-08-356-361-28	Sequence 28, Appl
62	125.5	16.7	76	2	US-08-769-967A-28	Sequence 28, Appl
63	124.5	16.6	181	2	US-08-640-977-3	Sequence 3, Appli
64	123.5	16.4	574	6	5378464-3	Patent No. 5378464
65	120	16.0	610	1	US-08-309-470-3	Sequence 3, Appli
66	120	16.0	610	3	US-09-209-668-89	Sequence 89, Appl
67	120	16.0	610	4	US-09-009-490A-89	Sequence 89, Appl
68	120	16.0	610	6	5217870-2	Patent No. 5217870
69	118	15.7	484	2	US-08-252-494C-9	Sequence 9, Appli
70	118	15.7	484	3	US-09-275-197-9	Sequence 9, Appli
71	115	15.3	126	6	5514582-45	Patent No. 5514582
72	114	15.2	830	1	US-08-111-158-4	Sequence 4, Appli
73	114	15.2	830	5	PCT-US91-05059-2	Sequence 2, Appli
74	112.5	15.0	830	6	5378464-2	Patent No. 5378464
75	111.5	14.8	240	3	US-08-824-692-23	Sequence 23, Appl
76	110.5	14.7	216	3	US-08-824-692-24	Sequence 24, Appl
77	110	14.6	123	3	US-08-824-692-37	Sequence 37, Appl
78	110	14.6	123	3	US-08-824-692-38	Sequence 38, Appl
79	109	14.5	127	6	5514582-2	Patent No. 5514582
80	108.5	14.4	128	6	5514582-4	Patent No. 5514582
81	106	14.1	62	1	US-08-213-266A-12	Sequence 12, Appl
82	106	14.1	62	1	US-08-388-675-12	Sequence 12, Appl
83	106	14.1	62	3	US-08-477-860C-12	Sequence 12, Appl
84	106	14.1	125	6	5514582-39	Patent No. 5514582
85	105	14.0	128	6	5514582-42	Patent No. 5514582
86	103.5	13.8	127	6	5514582-47	Patent No. 5514582
87	101	13.4	62	1	US-08-213-266A-13	Sequence 13, Appl
88	101	13.4	62	1	US-08-388-675-13	Sequence 13, Appl
89	101	13.4	62	3	US-08-477-860C-13	Sequence 13, Appl
90	97.5	13.0	177	3	US-08-824-692-30	Sequence 30, Appl
91	97.5	13.0	207	2	US-08-640-977-5	Sequence 5, Appli
92	84.5	11.3	145	2	US-08-640-977-6	Sequence 6, Appli
93	83	11.1	385	1	US-08-340-539A-2	Sequence 2, Appli
94	83	11.1	385	2	US-08-481-592B-2	Sequence 2, Appli
95	83	11.1	933	2	US-08-383-200-1	Sequence 1, Appli
96	83	11.1	933	5	PCT-US89-03837-1	Sequence 1, Appli
97	82.5	11.0	339	1	US-08-258-629A-4	Sequence 4, Appli
98	82.5	11.0	339	1	US-08-481-932-4	Sequence 4, Appli
99	82.5	11.0	339	1	US-08-482-260-4	Sequence 4, Appli
100	82.5	11.0	339	1	US-08-536-785-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
5256642-6
;Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLITIC AGENT, AND THE METHODS OF
;USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24 SEP 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26 SEP 1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03 APR 1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01 APR 1988
;SEQ ID NO:6:
; LENGTH: 1466
5256642-6

Query Match 29.2%; Score 219; DB 6; Length 1466;
Best Local Similarity 36.7%; Pred. No. 2.2e-15;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCQSPPPTILNGRISYYS-TPIAVGIVIVIVYSC SGIFRLIGEKSLICLIKDKV 52
|:| |:|:|:| :|| | ||| :| | | | | |:|:|:|: | :||
Db 1015 ISCEPPPTISNG-DFYSNNRISFHNGIVVTYQCHTGPDGFQLFELVGERSIYCTSKDDQ 1072

Qy 53 DGTWDKPAPKCEYPNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINESMNGNKSV 111
| | | | |:| || | | | :: | : : | |:| :| | | |:|
Db 1073 VGVWSSPPPRCISINK-CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 112 WCQANNMWGPTRLPTCVSV 130
| | | ||| :|| | |
Db 1130 QCQTNGRWGP-KLPHCSRV 1147

RESULT 2
5472939-6
;Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01-APR-1988
;SEQ ID NO:6:
; LENGTH: 1466
5472939-6

Query Match 29.2%; Score 219; DB 6; Length 1466;
Best Local Similarity 36.7%; Pred. No. 2.2e-15;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCQSPPPTILNGRISYYS-TPIAVGIVIVIVYSC SGIFRLIGEKSLICLIKDKV 52
|:| |:|:|:| :|| | ||| :| | | | | |:|:|:|: | :||
Db 1015 ISCEPPPTISNG-DFYSNNRISFHNGIVVTYQCHTGPDGFQLFELVGERSIYCTSKDDQ 1072

Qy 53 DGTWDKPAPKCEYPNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINESMNGNKSV 111
| | | | |:| || | | | :: | : : | |:| :| | | |:|
Db 1073 VGVWSSPPPRCISINK-CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 112 WCQANNMWGPTRLPTCVSV 130
| | | ||| :|| | |
Db 1130 QCQTNGRWGP-KLPHCSRV 1147

RESULT 3
5256642-5
;Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLITIC AGENT, AND THE METHODS OF
;USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01-APR-1988
;SEQ ID NO:5:
; LENGTH: 1537
5256642-5

Query Match 29.2%; Score 219; DB 6; Length 1537;
Best Local Similarity 36.7%; Pred. No. 2.4e-15;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCQSPPPTILNGRISYYS-TPIAVGIVIVIVYSC SGIFRLIGEKSLICLIKDKV 52
|:| |:|:|:| :|| | ||| :| | | | | |:|:|:|: | :||
Db 1015 ISCEPPPTISNG-DFYSNNRISFHNGIVVTYQCHTGPDGFQLFELVGERSIYCTSKDDQ 1072

Qy 53 DGTWDKPAPKCEYPNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINESMNGNKSV 111
| | | | |:| || | | | :: | : : | |:| :| | | |:|
Db 1073 VGVWSSPPPRCISINK-CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 112 WCQANNMWGPTRLPTCVSV 130
| | | ||| :|| | |
Db 1130 QCQTNGRWGP-KLPHCSRV 1147

RESULT 4
5472939-5
;Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989

```
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03 APR 1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:5;
;       LENGTH: 1537
5472939 5

  Query Match             29.2%;  Score 219;  DB 6;  Length 1537;
  Best Local Similarity   36.7%;  Pred. No. 2.4e 15;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS-  TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           ||| || | ||   :||   |    ||| | |         | |:|||:|: | :||
Db    1015 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1073 VGVWSSPPPRCISTNK--CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1130 QCQTNGRWGP-KLPHCSRV 1147


RESULT 5
5256642 10
;Patent No. 5256642
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLTIC AGENT, AND THE METHODS OF
;USE THEREOF
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/588,128
;       FILING DATE: 24-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:10:
;       LENGTH: 1847
5256642 10

  Query Match             29.2%;  Score 219;  DB 6;  Length 1847;
  Best Local Similarity   36.7%;  Pred. No. 3e 15;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           ||| || | ||   :||   |    ||| | |         | |:|||:|: | :||
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1637 QCQTNGRWGP-KLPHCSRV 1654


RESULT 6
5472939 10
;Patent No. 5472939
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/138,823
;       FILING DATE: 19-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 588,128
;       FILING DATE: 24-SEP-1990
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:10:
;       LENGTH: 2306
5472939-10

  Query Match             29.2%;  Score 219;  DB 6;  Length 1847;
  Best Local Similarity   36.7%;  Pred. No. 3e-15;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------ SGTFRLIGEKSLLCITKDKV 52
           ||| || | ||   :||   |    |||               | |:|||:|: | :||
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1637 QCQTNGRWGP-KLPHCSRV 1654


RESULT 7
5256642-2
;Patent No. 5256642
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLTIC AGENT, AND THE METHODS OF
;USE THEREOF
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/588,128
;       FILING DATE: 24-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:2:
;       LENGTH: 2039
5256642-2

  Query Match             29.2%;  Score 219;  DB 6;  Length 2039;
  Best Local Similarity   36.7%;  Pred. No. 3.4e 15;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC-      SGTFRLIGEKSLLCITKDKV 52
           ||| || | ||   :||   |    || : | |         | |:|||:|: | :||
Db    1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1575 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  |-| :|| |  |
```

```
Db    1642 QCQINGRWGP KIPHCSRV 1649


RESULT 8
5472039/2
 Patent No. 5472039
;      APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;TIP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;      TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
;      NUMBER OF SEQUENCES: 30
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/08/108,825
;        FILING DATE: 19-OCT-1993
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: 588,128
;        FILING DATE: 24-SEP-1990
;        APPLICATION NUMBER: 412,745
;        FILING DATE: 26-SEP-1989
;        APPLICATION NUMBER: 332,865
;        FILING DATE: 03-APR-1989
;        APPLICATION NUMBER: 176,542
;        FILING DATE: 01-APR-1988
 SEQ ID NO:2:
;      LENGTH: 2039
5472039/2

  Query Match             29.2%;  Score 219;  DB 6;  Length 2039;
  Best Local Similarity   30.7%;  Pred. No. 3.4e-15;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy        2 ISCGSPPPTLNGKISYY.   TIAVGTVTKYSC      SGTEKLVGEPSIECTEKGEW 72
            ||  || | ||   ||  |   |||: | |          | |:|||:| | :||
Db     1617 ISCEPPPTISNG  DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1674

Qy       73 DGTWDKPAPKCEYFNKYSSCPPPIVPGGYETNGSIC YRHGDSVTFACKTNFSMNGNKSV 132
             | | | |:|   ||   |  | |    :: |:   :   : : | |:  | | | :|
Db     1675 VGVWSSPPPRCISTNK   CTAPEVENAIRVPGNRSFFSLTEITRFRCQPGFVMVGSHTV 1734

Qy      133 WCQANNMWGPTRLPTCVSV 150
            ||  |  ||  ||| |  |
Db     1642 QCQINGRWGP KIPHCSRV 1649


RESULT 9
US-07-906-983-2
  Sequence 2, Application US/07/906983
  Patent No. 5102268
  GENERAL INFORMATION:
    APPLICANT: Kotwal, Girish
    APPLICANT: Moss, Bernard
    TITLE OF INVENTION: Synthetic, Anti-Complement Protein and
    TITLE OF INVENTION: the Gene Encoding Same
    NUMBER OF SEQUENCES: 4
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Townsend and Townsend
      STREET: One Market Plaza, Steuart Tower, Suite 2000
      CITY: San Francisco
      STATE: California
      COUNTRY: USA
      ZIP: 94105
    COMPUTER READABLE FORM
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/07/906,983
      FILING DATE: 19920701
      CLASSIFICATION: 530
    ATTORNEY/AGENT INFORMATION:
      NAME: Weber, Kenneth A.
      REGISTRATION NUMBER: 31,677
      REFERENCE/DOCKET NUMBER: 15280-9
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 415 543-9600
      TELEFAX: 415 543-5043
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 263 amino acids
      TYPE: AMINO ACID
      TOPOLOGY: linear
    MOLECULE TYPE: protein
US-07-906-983-2

  Query Match             28.5%;  Score 214;  DB 1;  Length 263;
  Best Local Similarity   35.9%;  Pred. No. 8.3e-16;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy        2 ISCGSPPPTLNGKISYYSITIAVGTVTKYSCSGTEKLVGEPSIECTEKGVWTQKPAP 61
            :  ||| | ||| : |       |:|: |||: : |||  :||       | | | |
Db      146 VKPCQTPTSTCNGRNGYIEDFYIDGSVVIYSCNSSYSLLGNSSVIC  SGGEWSDPP 199

Qy       62 KCPYPNFYSS-TELPIVRYFIFI-SIYSHGIVTIAVEKNKMNYKVWTQANNMWG 120
            :         || | :  ||   |  | : |: | ||  : ::::: | |  | |
Db      200 TCQTV----KCPHPTISNGYLSSGEKRSYSYNTNVTFKCKYGYRKVSSSSSTCSPGNTWK 255

Qy      121 PTRLTCV 128
            |  || ||
Db      256 PELPKCV 262


RESULT 10
US-08-356-361-31
; Sequence 31, Application US/08356361
; Patent No. 5833989
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I.
;    TITLE OF INVENTION:  No. 5833989el Compounds
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,361
;      FILING DATE:  03-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H.
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P30423
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5019
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  133 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-356-361-31
```

```
  Query Match             26.4%;  Score 198;  DB 2;  Length 133;
  Best Local Similarity   38.1%;  Pred. No. 1.9e-14;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSTPIAVGTVIRYSC---SG---TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db       8 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 62

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | | |  :  : : :   : | | |:  | | | : | |
Db      63 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 119

Qy     114 QANNMWGPTRLPTC 127
           || | | |  ||:|
Db     120 QALNKWEP-ELPSC 132


RESULT 11
US-08-769-967A-41
; Sequence 41, Application US/08769967A
; Patent No. 5859223
;  GENERAL INFORMATION:
;    APPLICANT:  Mossakowska, Danuta E.I.
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman, Anne Mary
;    TITLE OF INVENTION:  Soluble CR1 Derivatives
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,967A
;      FILING DATE:
;      CLASSIFICATION:  546
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/440,569
;      FILING DATE:  15-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, William T.
;      REGISTRATION NUMBER:  30,954
;      REFERENCE/DOCKET NUMBER:  P30423C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5464
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  133 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-769-967A-41

  Query Match             26.4%;  Score 198;  DB 2;  Length 133;
  Best Local Similarity   38.1%;  Pred. No. 1.9e-14;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSTPIAVGIVIRYSC---SG---TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db       8 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 62

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | | |  :  : : :   : | | |:  | | | : | |
Db      63 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 119

Qy     114 QANNMWGPTRLPTC 127
           || | | |  ||:|
Db     120 QALNKWEP-ELPSC 132


RESULT 12
US-08-356-361-29
; Sequence 29, Application US/08356361
; Patent No. 5833989
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I
;    TITLE OF INVENTION:  No. 5833989el Compounds
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham-Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,361
;      FILING DATE:  03-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H.
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P30423
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5019
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  254 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-356-361-29

  Query Match             26.4%;  Score 198;  DB 2;  Length 254;
  Best Local Similarity   38.1%;  Pred. No. 4.5e-14;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSTPIAVGIVIRYSC---SG---TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db     124 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 184

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | | |  :  : : :   : | | |:  | | | : | |
Db     184 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTC 127
           || | |    ||:|
Db     241 QALNKWEP-ELPSC 253


RESULT 13
US-08-769-967A-29
; Sequence 29, Application US/08769967A
; Patent No. 5859223
;  GENERAL INFORMATION:
;    APPLICANT:  Mossakowska, Danuta E.I
```

```
;       APPLICANT:  Smith, Richard A.G.
;       APPLICANT:  Dodd, Ian
;       APPLICANT:  Freeman, Anne Mary
;       TITLE OF INVENTION:  Soluble CRI Derivatives
;       NUMBER OF SEQUENCES:  33
;       CORRESPONDENCE ADDRESS:
;         ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;         STREET:  P.O. Box 1539
;         CITY:  King of Prussia
;         STATE:  Pennsylvania
;         COUNTRY:  USA
;         ZIP:  19406
;       COMPUTER READABLE FORM:
;         MEDIUM TYPE:  Floppy disk
;         COMPUTER:  IBM PC compatible
;         OPERATING SYSTEM:  PC DOS/MS DOS
;         SOFTWARE:  PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/769,967A
;         FILING DATE:
;         CLASSIFICATION:  536
;       PRIOR APPLICATION DATA:
;         APPLICATION NUMBER:  08/440,569
;         FILING DATE:  15-May-1995
;       ATTORNEY/AGENT INFORMATION:
;         NAME:  King, William I.
;         REGISTRATION NUMBER:  30,924
;         REFERENCE/DOCKET NUMBER:  P30423C2
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE:  (610) 270-5064
;         TELEFAX:  (610) 270-5090
;     INFORMATION FOR SEQ ID NO:  29:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH:  254 amino acids
;         TYPE:  amino acid
;         TOPOLOGY:  linear
;       MOLECULE TYPE:  peptide
;       FRAGMENT TYPE:  N-terminal
US-08-769-967A-29

  Query Match              26.4%;  Score 198;  DB 2;  Length 254;
  Best Local Similarity    38.1%;  Pred. No. 4.5e-14;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPEILNGRISYYSTPIAVGTVIRYSC   SG-  TPRLIGEKSLLLTIKDKVDG 54
           | || || | ||  ||         |:|: | |    ||    | |:|| : : . : |   |
Db     124 IPCGLPPELINGDFTISNPFNEHYGISVVIYKCNDGSGGRKVFEEAVGIPSTYCISNDDQVG 183

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRPGDSVTFACKTNFSMNGNKSVWC 113
           |   |:::  ||   | || | : : : :   : | | |: | | | | : | |
Db     184 IWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTC 127
           || | | | . |:|
Db     241 QALNKWEP ELPSC 253


RESULT 14
US-08-456-461-30
; Sequence 30, Application US/08456461
; Patent No. 5833989
; GENERAL INFORMATION:
;     APPLICANT:  Smith, Richard A.G.
;     APPLICANT:  Dodd, Ian
;     APPLICANT:  Freeman Mary A.
;     APPLICANT:  Mossakowska, Danuta E.I.
;     TITLE OF INVENTION:  N- 5833989el Compounds
;     NUMBER OF SEQUENCES:  41
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/456,461
;       FILING DATE:  03-Jul-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jervis, Herbert B.
;       REGISTRATION NUMBER:  41,171
;       REFERENCE/DOCKET NUMBER:  P30423
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5019
;       TELEFAX:  (610) 270-5090
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  N-terminal
US-08-456-461-30

  Query Match              26.0%;  Score 195;  DB 2;  Length 254;
  Best Local Similarity    38.1%;  Pred. No. 9.6e-14;
  Matches   51;  Conservative   14;  Mismatches   58;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPEILNGRISYYSTPIAVGTVIRYSC   SG-  TPRLIGEKSLLLTIKDKVDG 54
           | || || | ||  ||         |:|: | |    ||    | |:|| : : . : |   |
Db     124 IPCGLPPELINGDFTISNPFNEHYGISVVIYKCNDSGGRKVFEEAVGIPSTYCISNDDQVG 183

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRPGDSVTFACKTNFSMNGNKSVWC 113
           |   |:::  ||   | || | : : : :   : | | |: | | | | : | |
Db     184 IWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTC 127
           || | | | . |:|
Db     241 QALNKWEP ELPSC 253


RESULT 15
US-08-769-967A-30
; Sequence 30, Application US/08769967A
; Patent No. 5859223
; GENERAL INFORMATION:
;     APPLICANT:  Mossakowska, Danuta E.I.
;     APPLICANT:  Smith, Richard A.G.
;     APPLICANT:  Dodd, Ian
;     APPLICANT:  Freeman, Anne Mary
;     TITLE OF INVENTION:  Soluble CRI Derivatives
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/769,967A
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/440,569
```

```
;       FILING DATE:  15-May-1995
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  King, William T.
;       REGISTRATION NUMBER:  30,954
;       REFERENCE/DOCKET NUMBER:  P3042C2
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5464
;       TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-769-967A-30

  Query Match             26.0%;  Score 195;  DB 2;  Length 254;
  Best Local Similarity   38.1%;  Pred. No. 9.6e-14;
  Matches   51;  Conservative   13;  Mismatches   58;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPP LNGR ISYYSTPIAVGTVIRYSC- -SG---TFRLIGEKSLLCITKDKVDG 54
           | || || - ||  ||        |:: | |   ||    | |:|| |: | : |   |
Db     124 IPCGLPPTIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 183

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | |  |  : : : :   : | | |:  | | |   | |
Db     184 IWSGPAPQCIIPNK---CTIPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPHRVKC 240

Qy     114 QANNMWGPTRLPTC 127
           || | | |  ||:|
Db     241 QALNKWEP ELPSC 253


RESULT 16
US-08-981-234B-2
; Sequence 2, Application US/08981234B
; Patent No. 6270997
; GENERAL INFORMATION:
;    APPLICANT:  TOYOMURA, KOJI
;    APPLICANT:  MURAKAMI, HIROSHI
;    APPLICANT:  SHIGEHISA, TAMOTSU
;    TITLE OF INVENTION:  DNA ENCODING A PORCINE COMPLEMENT
;    TITLE OF INVENTION:  INHIBITOR
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;       STREET:  PO BOX 747
;       CITY:  FALLS CHURCH
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/981,234B
;       FILING DATE:  12-DEC-1997
;       CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  MURPHY JR, GERALD M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  2520-111P
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703 205 8000
;       TELEFAX:  703 205 8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  363 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-981-234B-2

  Query Match             25.3%;  Score 190;  DB 4;  Length 363;
  Best Local Similarity   33.6%;  Pred. No. 5.4e-13;
  Matches   45;  Conservative   20;  Mismatches   94;  Indels   16;  Gaps    7;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC-SGT----FRLIGEKSLLCIIKDKVDGT 55
           | |  |  | ||: : :          |: ||| | |     | :|| || || ||:
Db     169 ILCKPPQEIPNGKYTNSHKDVFEYNEVVTYSCLSSTGPDEFSLVGESSLFCICKDE- - 224

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG STEYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           |    |:|    |   || |:|| |  : |  : : :   | | |   |:::   :: |
Db     225 WSSDPPEC  --KVVKCPYPVVPNGEIVSCPGSREYYKAEVVFKCNAGFTLHGRDTIVCG 280

Qy     115 ANNMWGPTRLPTCV 128
           ||: | |  :| |:
Db     281 ANSTWEP EMPQCI 293


RESULT 17
US-08-139-195-2
; Sequence 2, Application US/08139195
; Patent No. 6218520
; GENERAL INFORMATION:
;    APPLICANT:  Atkinson, John P.
;    TITLE OF INVENTION:  RECOMBINANTLY PRODUCED HUMAN MEMBRANE
;    TITLE OF INVENTION:  COFACTOR PROTEIN (MCP)
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Patrea L. Pabst
;       STREET:  1100 Peachtree Street, Suite 2800
;       CITY:  Atlanta
;       STATE:  Georgia
;       COUNTRY:  USA
;       ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/139,195
;       FILING DATE:  20-OCT-1993
;       CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/948,350
;       FILING DATE:  21-SEP-1992
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/384,213
;       FILING DATE:  21-JUL-1989
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Pabst, Patrea L.
;       REGISTRATION NUMBER:  31,284
;       REFERENCE/DOCKET NUMBER:  WUL00100N
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (404)-815-6508
;       TELEFAX:  (404)-815-6555
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  384 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;       ORGANISM:  Human Membrane CoFactor Protein (MCP)
US-08-139-195-2

  Query Match             21.3%;  Score 160;  DB 4;  Length 384;
```

```
  Best Local Similarity   27.9%;  Prod. No. 1;le 09;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISGSPPPIINCRISYYSTPI AVGTVIRYSCS    GTPRLIGEKSLLCIIKDKVDGT 55
           : |  || | ||: ::    :       : ||         | |||| :: |      :
Db     160 VLCTIPPKIKNGKHIESEVEVFEEYLDAVIYSCDPAPGPDPPSLIGESTIYC    GDNSV 215

Qy      56 WDKPAFKCEYENKYSSCSEECIVIGGYKIRC SIPYKHCDSVFEACKINFSMNCNKSVWCQ 114
           | : ||:|    |  |  |:|  | :| |    : :  :| | |   | :::: :: |
Db     216 WSRAAPEC    KVVKCPFPVVENGKQISGFGKKFYYKATVMFECDKGFYLIGSDTIVCD 271

Qy     115 ANNMWGPTRLPICVSV 130
           :|: | |  :| | :
Db     272 SNSIWDPP VPKCLKV 286


RESULT 18
US-08-310-416A-18
; Sequence 18, Application US/08310416A
; Patent No. 5679546
;  GENERAL INFORMATION:
;    APPLICANT:  Jone Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 80 386DX
;      OPERATING SYSTEM:  MS DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,416A
;      FILING DATE:  22-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul J. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  06150/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  169 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-310-416A-18

  Query Match              21.2%;  Score 159;  DB 1;  Length 169;
  Best Local Similarity   27.9%;  Prod. No. 5e-10;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISGSPPPIINCRISYYSTPI AVGTVIRYSCS    GTPRLIGEKSLLCIIKDKVDGT 55
           : |  || | ||: ::    :       : ||         | |||| :: |      :
Db       2 VLCTIPPKIKNGKHIESEVEVFEEYLDAVIYSCDPAPGPDPPSLIGESTIYC    GDNSV 57

Qy      56 WDKPAFKCEYENKYSSCSEECIVIGGYKIRC SIPYKHCDSVFEACKINFSMNCNKSVWCQ 114
           | : ||:|    |  |  |:|  | :| |    : :  :| | |   | :::: :: |
Db      58 WSRAAPEC    KVVKCPFPVVENGKQISGFGKKFYYKATVMFECDKGFYLIGSDTIVCD 113

Qy     115 ANNMWGPTRLPICVSV 130
           :|: | |  :| | :
Db     114 SNSIWDPP VPKCLKV 128


RESULT 19
US-08-888-171-18
; Sequence 18, Application US/08888171
; Patent No. 5851528
;  GENERAL INFORMATION:
;    APPLICANT:  Jone Long, Ko
;    APPLICANT:  Higgins, Paul J.
;    APPLICANT:  Yeh, C. Grace
;    TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION:  ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,171
;      FILING DATE:  03-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/310,416
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06150/005002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  169 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-888-171-18

  Query Match              21.2%;  Score 159;  DB 2;  Length 169;
  Best Local Similarity   27.9%;  Prod. No. 5e-10;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISGSPPPIINCRISYYSTPI AVGTVIRYSCS    GTPRLIGEKSLLCIIKDKVDGT 55
           : |  || | ||: ::    :       : ||         | |||| :: |      :
Db       2 VLCTIPPKIKNGKHIESEVEVFEEYLDAVIYSCDPAPGPDPPSLIGESTIYC    GDNSV 57

Qy      56 WDKPAFKCEYENKYSSCSEECIVIGGYKIRC SIPYKHCDSVFEACKINFSMNCNKSVWCQ 114
           | : ||:|    |  |  |:|  | :| |    : :  :| | |   | :::: :: |
Db      58 WSRAAPEC    KVVKCPFPVVENGKQISGFGKKFYYKATVMFECDKGFYLIGSDTIVCD 113

Qy     115 ANNMWGPTRLPICVSV 130
           :|: | |  :| | :
Db     114 SNSIWDPP VPKCLKV 128


RESULT 20
US-08-310-416A-13
; Sequence 13, Application US/08310416A
; Patent No. 5679546
;  GENERAL INFORMATION:
;    APPLICANT:  Jone Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  MS DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/310,416A
;       FILING DATE:  22-SEP-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Paul T. Clark
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  06180/005001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
;     MOLECULE TYPE  protein
US-08-310-416A-13

  Query Match             21.2%;  Score 159;  DB 1;  Length 254;
  Best Local Similarity   27.9%;  Pred. No. 8.6e-10;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI AVGTVIRYSCS----GTFRLIGEKSLLCITKDKVDGT 55
           : |  || | ||: ::    :      : |||        | |||| :: |      :
Db     126 VLCTPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDPFSLIGESTIYC----GDNSV 181

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | : ||:|    |   | |:| | :| |    : :  :| | |   | ::|: :: |
Db     182 WSRAAPEC--KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 237

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: | |  :| |: |
Db     238 SNSTWDPP-VPKCLKV 252


RESULT 21
US-08-888-171-13
; Sequence 13, Application US/08888171
; Patent No. 5851528
;  GENERAL INFORMATION:
;     APPLICANT:  Jone Long, Ko
;     APPLICANT:  Higgins, Paul J.
;     APPLICANT:  Yeh, C. Grace
;     TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;     TITLE OF INVENTION:  ACTIVATION
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson, P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/888,171
;       FILING DATE:  03-JUL-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/310,416
;       FILING DATE:  22-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Freeman, John W.
;       REGISTRATION NUMBER:  29,066
;       REFERENCE/DOCKET NUMBER:  06180/005002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-507
;       TELEFAX:  617/542-890
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE  protein
US-08-888-171-13

  Query Match             21.2%;  Score 159;  DB 2;  Length 254;
  Best Local Similarity   27.9%;  Pred. No. 8.6e-10;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS----GTFRLIGEKSLLCITKDKVDGT 55
           : |  || | ||: ::    :      : |||        | |||| :: |      :
Db     126 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC----GDNSV 181

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | : ||:|    |   | |:| | :| |    : :  :| | |   | ::|: :: |
Db     182 WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 237

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: | |  :| |: |
Db     238 SNSTWDPP-VPKCLKV 252


RESULT 22
US-08-435-149-1
; Sequence 1, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;     APPLICANT:  INNIS, MICHAEL A.
;     APPLICANT:  ZAKOR, ISABEL
;     APPLICANT:  CREASEY, ABLA A.
;     TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;     TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;     NUMBER OF SEQUENCES:  26
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CHIRON CORPORATION
;       STREET:  INTELLECTUAL PROPERTY - 8440, P.O. BOX 8097
;       CITY:  EMERYVILLE
;       STATE:  CALIFORNIA
;       COUNTRY:  U.S.A.
;       ZIP:  94662-8097
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/435,149
;       FILING DATE:  05-MAY-1995
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SAVEREIDE, PAUL B.
;       REGISTRATION NUMBER:  36,914
;       REFERENCE/DOCKET NUMBER:  0989.101
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 601-2585
;       TELEFAX:  (510) 655-3542
```

```
;       TELEX:  N/A
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-435-149-1

  Query Match              21.2%;  Score 159;  DB 2;  Length 254;
  Best Local Similarity    27.9%;  Pred. No. 8.6e-10;
  Matches   88;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGIVIERYSCS    -GIERLIGEKSLICLIKDKVDGI 55
           : |  || | ||: ::    :      : |||       | |||| :: |     :
Db     126 VICLIPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDDPESLIGESIIYC    GDNSV 181

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-SIPYRHGDSVIFACKINFSMNGNKSVWCQ 114
           | : ||:|    |   |  |::  : :|  |    :  :  :| | |   | :::|  :: |
Db     182 WSRAAPEC    KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDEIVCD 237

Qy     115 ANNMWGPTRLPTCVSV 130
           :|  | |  :| |: |
Db     238 SNSIWDPP VPKCLKV 252


RESULT 23
US-08-416A-16
; Sequence 16, Application US/08416A
; Patent No. 5677456
;  GENERAL INFORMATION:
;    APPLICANT:  Jone-Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50 or 55 SX
;      OPERATING SYSTEM:  MS DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/416,416A
;      FILING DATE:  22-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul F. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  07180/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  293 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-416A-16

  Query Match              21.2%;  Score 159;  DB 1;  Length 293;
  Best Local Similarity    27.9%;  Pred. No. 1e-09;
  Matches   88;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGIVIERYSCS    -GIERLIGEKSLICLIKDKVDGI 55
           : |  || | ||: ::    :      : |||       | |||| :: |     :
Db     126 VICLIPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDDPESLIGESIIYC    GDNSV 181

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-SIPYRHGDSVIFACKINFSMNGNKSVWCQ 114
           | : ||:|    |   |  |::  : :|  |    :  :  :| | |   | :::|  :: |
Db     182 WSRAAPEC    KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDEIVCD 237

Qy     115 ANNMWGPTRLPTCVSV 130
           :|  | |  :| |: |
Db     238 SNSIWDPP VPKCLKV 252


RESULT 24
US-08-888-171-16
; Sequence 16, Application US/08888171
; Patent No. 5851528
;  GENERAL INFORMATION:
;    APPLICANT:  Jone-Long, Ko
;    APPLICANT:  Higgins, Paul J.
;    APPLICANT:  Yeh, C. Grace
;    TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION:  ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,171
;      FILING DATE:  03-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/310,416
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  07180/005002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  293 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-888-171-16

  Query Match              21.2%;  Score 159;  DB 2;  Length 293;
  Best Local Similarity    27.9%;  Pred. No. 1e-09;
  Matches   88;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGIVIERYSCS    -GIERLIGEKSLICLIKDKVDGI 55
           : |  || | ||: ::    :      : |||       | |||| :: |     :
Db     126 VICLIPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDDPESLIGESIIYC    GDNSV 181

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-SIPYRHGDSVIFACKINFSMNGNKSVWCQ 114
           | : ||:|    |   |  |::  : :|  |    :  :  :| | |   | :::|  :: |
Db     182 WSRAAPEC -  KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDEIVCD 237

Qy     115 ANNMWGPTRLPTCVSV 130
           :|  | |  :| |: |
Db     238 SNSTWDPP VPKCLKV 252
```

```
RESULT 25
US-08-528-057-42
; Sequence 42, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,057
;      FILING DATE:  CONCURRENTLY HEREWITH
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,686
;      FILING DATE:  11-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU91/00199
;      FILING DATE:  10-MAY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK0133/90
;      FILING DATE:  11-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  17227/112 DACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  370 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-528-057-42

  Query Match             21.2%;  Score 159;  DB 2,  Length 370,
  Best Local Similarity   27.9%;  Pred. No. 1.4e-09;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPFILNGRISYYSTPI-AVGTVIRYSCS----GTFRLIGEKSLLCITKDKVDGT 55
           : |  || | ||: ::    :       : |||       | |||| :: |      :
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC----GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPPPIVPGGYKIRG-SIPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
          | : ||:    |  |  |:|  | :| |    : :  :| | |   | ::|: :: |
Db     216 WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: | |  :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 26
US-08-528-057-44
; Sequence 44, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,057
;      FILING DATE:  CONCURRENTLY HEREWITH
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,686
;      FILING DATE:  11-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU91/00199
;      FILING DATE:  10-MAY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK0133/90
;      FILING DATE:  11-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  17227/112 DACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  373 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-528-057-44

  Query Match             21.2%;  Score 159;  DB 2;  Length 373;
  Best Local Similarity   27.9%;  Pred. No. 1.4e-09;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPFILNGRISYYSTPI-AVGTVIRYSCS----GTFRLIGEKSLLCITKDKVDGT 55
           : |  || | ||: ::    :       : |||       | |||| :: |      :
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC----GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPPPIVPGGYKIRG-SIPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
          | : ||:    |  |  |:|  | :| |    : :  :| | |   | ::|: :: |
Db     216 WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: | |  :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 27
US-08-528-057-2
; Sequence 2, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
```

```
      CORRESPONDENCE ADDRESS:
        ADDRESSEE:  Foley & Lardner
        STREET:  3000 K Street, N.W., Suite 500
        CITY:  Washington
        STATE:  D.C.
        COUNTRY:  USA
        ZIP:  20007-5109
      COMPUTER READABLE FORM
        MEDIUM TYPE:  Floppy disk
        COMPUTER:  IBM PC Compatible
        OPERATING SYSTEM:  PC DOS/MS DOS
        SOFTWARE:  PatentIn Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  US/08/528,057
        FILING DATE:  CONCURRENTLY HEREWITH
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  US 07/961,686
        FILING DATE:  11 JAN 1993
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  PCT/AU91/00199
        FILING DATE:  10 MAY 1991
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  AU PK0133/90
        FILING DATE:  11 MAY 1990
      ATTORNEY/AGENT INFORMATION:
        NAME:  BENT, Stephen A.
        REGISTRATION NUMBER:  29,768
        REFERENCE/DOCKET NUMBER:  37227/212 DAC5
      TELECOMMUNICATION INFORMATION:
        TELEPHONE:  (202)672-5300
        TELEFAX:  (202)672-5399
        TELEX:  904136
    INFORMATION FOR SEQ ID NO:  2:
      SEQUENCE CHARACTERISTICS:
        LENGTH:  377 amino acids
        TYPE:  amino acid
        TOPOLOGY:  linear
      MOLECULE TYPE:  protein
US-08-528-057-2

  Query Match             21.2%;  Score 159;  DB 2;  Length 377;
  Best Local Similarity   27.9%;  Pred. No. 1.4e-09;
  Matches   48;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSEPPFILNGRISYYSTPI-AVGTVIRYSCS      GTPFRLIGEKSLLCITKDKVDGT 55
           :|  | | :|:  ::     :  ||||          | |||| : |      :
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-  -GDNSV 215

Qy       56 WDKPAPKCEYFNKYSSCPEPTVPGGYKIRG-STPYRDGDSVTFACKTNFSMNGNKSVWCQ 114
            | : ||:|     |   |  |:|  | :| |     : :  :| |      | :::: : |
Db     216 WSRAAPEC--KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDPGFYLIGSDTIVCD 271

Qy      115 ANNMWGPTRLPTCVSV 130
            :|: | |  :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 29
US-08-435-149-3
; Sequence 3, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;    APPLICANT:  INNIS, MICHAEL A.
;    APPLICANT:  ZAROR, ISABEL
;    APPLICANT:  CREASEY, ABLA A.
;    TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;    TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  INTELLECTUAL PROPERTY  R440, P.O. BOX 8097
;      CITY:  EMERYVILLE
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,149
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAVEREIDE, PAUL B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0989.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2585
;      TELEFAX:  (510) 655-3542
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  577 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-435-149-3

  Query Match             21.2%;  Score 159;  DB 2;  Length 577;
  Best Local Similarity   33.3%;  Pred. No. 2.5e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy       3 SCGSEPPFILNGRISYYSTP-IAVGTVIRYSCSGTFPLIGEKSLLCITKDKVDGT  WD 57
           || :|  | ||:|     |  |  |  | :||:  ::| |  |  |:     : |:   |
Db     382 SCPNIGEIRNGQI---DVPGGILFGATISFSCNTGYKLFGSISSFCL-  ISGSSVQWS 434

Qy      58 KPAPKCEYFNKYSSCPEPTVPGGYKIRGSTPYRDGDSVTFACKTNFSMNGNKSVWCQAN 116
             | :|    :  || |      :|     | :  ||||||  :: | |  |:::  |
Db     435 DPLPEC----REIYCPAPPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 490
```

```
Qy     117 N   MW GPTRLPTC 127
           |    | ||   | |
Db     491 NDEGEWSGPP--PEC 503


RESULT 30
US-09-475-460A-32
; Sequence 32, Application US/09475460A
; Patent No. 6316253
; GENERAL INFORMATION:
;  APPLICANT:  Scott, Elizabeth
;  APPLICANT:  Innis, Michael
;  TITLE OF INVENTION:  EXPRESSION VECTORS, TRANSFECTION SYSTEMS, AND METHOD OF
;  TITLE OF INVENTION:  USE THEREOF
;  FILE REFERENCE:  1527.003
;  CURRENT APPLICATION NUMBER: US/09/475,460A
;  CURRENT FILING DATE:  1999-12-30
;  NUMBER OF SEQ ID NOS:  32
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 32
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: CAB2
US-09-475-460A-32

  Query Match             21.2%;  Score 159;  DB 4;  Length 611;
  Best Local Similarity   33.3%;  Pred. No. 2.7e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy       3 SCGSPPPILNGRISYYSTP--IAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT---WD 57
           || :|  | ||:|     |  |  |  | :||:  ::| |  |  |:    : |:   |
Db     416 SCPNPGEIRNGQI---DVPGGILFGATISFSCNTGYKLFGSTSSFCL----ISGSSVQWS 468

Qy      58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSVWCQAN 116
             | |:|    :   || |       |:|    | :  |||:||   |:| |  |::|  |
Db     469 DPLPEC---REIYCPAPPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 524

Qy     117 N   MW GPTRLPTC 127
           |    | ||   | |
Db     525 NDEGEWSGPP--PEC 537


RESULT 31
US-08-435-149-2
; Sequence 2, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;    APPLICANT:  INNIS, MICHAEL A.
;    APPLICANT:  ZAROR, ISABEL
;    APPLICANT:  CREASEY, ABLA A.
;    TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;    TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  INTELLECTUAL PROPERTY  R440, P.O. BOX 8097
;      CITY:  EMERYVILLE
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,149
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  540
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAVEREIDE, PAUL B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0989.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2585
;      TELEFAX:  (510) 655-3542
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-435-149-2

  Query Match             21.1%;  Score 158;  DB 2;  Length 323;
  Best Local Similarity   33.3%;  Pred. No. 1.3e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy       3 SCGSPPPILNGRISYYSTP--IAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT   WD 57
           || :|  | ||:|     |  |  |  | :||:  ::| |  |  |:    : |:   |
Db     128 SCPNPGEIRNGQI---DVPGGILFGATISFSCN GYKLFGSTSSFCL----ISGSSVQWS 180

Qy      58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSVWCQAN 116
             | |:|    :   || |       |:|    | :  |||:||   |:| |  |::|  |
Db     181 DPLPEC---REIYCPAPPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 236

Qy     117 N---MW GPTRLPTC 127
           |    | ||   | |
Db     237 NDEGEWSGPP--PEC 249


RESULT 32
US-08-310-416A-14
; Sequence 14, Application US/08310416A
; Patent No. 5673546
;  GENERAL INFORMATION:
;    APPLICANT:  Jone-Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,416A
;      FILING DATE:  22-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  06180/035001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  324 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-310-416A-14
```

```
  Query Match             21.1%;  Score 158.5;  DB 1;  Length 424;
  Best Local Similarity   34.3%;  Pred. No. 1.3e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy       3 SCGSPPPILNGRISYYSTP-IAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT 55
           ||  |  | ||:|     | |  |  | :||:  ::| |  |  |:    : |:   |
Db     128 SCPNIGETRNGQI-DVRGGILFGATISFSCNTGYKLFGSTSSFCL--ISGSSVQWS 180

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQAN 115
           | |:|    :  || |     |:|   | :  |||:||   |:| |  |:::  |
Db     181 DPLPEC    RETYCPAPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 236

Qy     116 N  MW GPTRLPTCV 128
           |   | ||   | |
Db     237 NDEGEWSGPP-PEC 249


RESULT 34
US-08-888-171-14
; Sequence 34, Application US/08888171
; Patent No. 5861528
;  GENERAL INFORMATION:
;    APPLICANT:  Jane Lebl, Ke
;    APPLICANT:  Higgins, Paul J.
;    APPLICANT:  Yeh, C. Grace
;    TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION:  ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,171
;      FILING DATE:  04-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/410,416
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,056
;      REFERENCE/DOCKET NUMBER:  06180/005002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  424 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-888-171-14

  Query Match             21.1%;  Score 158.5;  DB 2;  Length 424;
  Best Local Similarity   34.3%;  Pred. No. 1.3e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy       3 SCGSPPPILNGRISYYSTP-IAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT 55
           ||  |  | ||:|     | |  |  | :||:  ::| |  |  |:    : |:   |
Db     128 SCPNIGETRNGQI-DVRGGILFGATISFSCNTGYKLFGSTSSFCL--ISGSSVQWS 180

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQAN 115
           | |:|    :  || |     |:|   | :  |||:||   |:| |  |:::  |
Db     181 DPLPEC    RETYCPAPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 236

Qy     116 N  MW GPTRLPTCV 128
           |   | ||   | |
Db     237 NDEGEWSGPP-PEC 249


RESULT 34
US-08-528-057-46
; Sequence 46, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,057
;      FILING DATE:  CONCURRENTLY HEREWITH
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,686
;      FILING DATE:  11-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU91/00199
;      FILING DATE:  10-MAY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK0133/90
;      FILING DATE:  11-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  37227/212 DACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904146
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  424 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-528-057-46

  Query Match             20.6%;  Score 155;  DB 2;  Length 424;
  Best Local Similarity   27.6%;  Pred. No. 3.2e-09;
  Matches   37;  Conservative   25;  Mismatches   56;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS-   -GTFRLIG-KSLLCITKDKVDGT 55
           : |  || | ||: ::    :      : |||         | |||| :: |      :
Db     160 VLCTPPPKIKNGKHIFSEVEVFEYLDAVIYSCDPAPGPDPFSLIGESTIYC-   GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG SIPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | : |||:|    |  | |:|  | :| |    : :  :| | |    | ::|: :: |
Db     216 WSKAAPEC----KVVKCKFPVVENGNQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy     115 ANNMWGPTRLPTCV 128
           :|: | |  :| |:
Db     272 SNSTWDPP VPKCL 284
```

```
RESULT 35
US-08-456-361-27
; Sequence 27, Application US/08456361
; Patent No. 5834989
; GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I.
;    TITLE OF INVENTION:  No. 5834989el Compounds
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham-Corporate Intellecutal Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,361
;      FILING DATE:  03-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H.
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P30423
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5019
;      TELEFAX:  (610) 270 5090
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  197 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-456-361-27

  Query Match             20.4%;  Score 153.5;  DB 2;  Length 197;
  Best Local Similarity   30.4%;  Pred. No. 2.5e-09;
  Matches   42;  Conservative   21;  Mismatches   56;  Indels   19;  Gaps    8;

Qy       3 SCGSPPPILNGRISYYSTPIAVGIVIRYSCSGTFRLIGEKSLLC-ITKDKVDGTWDKPAP 61
           || :||  :|| : :    |  |: |:|||:  :|||  |  | |: | |   ||   |
Db      63 SCRNPPDPVNGMV-HVIKGIQFGSQIKYSCTKGYRLIGSSSATCIISGDTV--IWDNETP 119

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTN------FSMNGNKSVWCQA 115
             |:       |  |  | :      : :|  ||: |        | : |  |::| :
Db     120 ICDRIP--CGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTS 177

Qy     116 NN----MW-GPTRLPTCV 128
           |:    :| ||   | |:
Db     178 NDDQVGIWSGPA--PQCI 193


RESULT 36
US-08-769-967A-27
; Sequence 27, Application US/08769967A
; Patent No. 5859223
; GENERAL INFORMATION:
;    APPLICANT:  Mossakowska, Danuta E.I.
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman, Anne Mary
;    TITLE OF INVENTION:  Soluble CR1 Derivatives
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,967A
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/440,569
;      FILING DATE:  15-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, William T.
;      REGISTRATION NUMBER:  30,954
;      REFERENCE/DOCKET NUMBER:  P30423C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5364
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  197 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-769-967A-27

  Query Match             20.4%;  Score 153.5;  DB 2;  Length 197;
  Best Local Similarity   30.4%;  Pred. No. 2.5e-09;
  Matches   42;  Conservative   21;  Mismatches   56;  Indels   19;  Gaps    8;

Qy       3 SCGSPPPILNGRISYYSTPIAVGIVIRYSCSGTIRLIGEKSLLC-ITKDKVDGTWDKPAP 61
           || :||  :|| : :    |  |: |:|||:  :|||  |  | |: | |   ||   |
Db      63 SCRNPEDPVNGMV-HVIKGIQFGSQIKYSCIKGYRLIGSSSATCIISGDTV--IWDNETP 119

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTN------FSMNGNKSVWCQA 115
             |:       |  |  | :      : :|  ||: |        | : |  |::| :
Db     120 ICDRIP--CGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTS 177

Qy     116 NN----MW-GPTRLPTCV 128
           |:    :| ||   | |:
Db     178 NDDQVGIWSGPA--PQCI 193


RESULT 37
US-08-824-692-32
; Sequence 32, Application US/08824692
; Patent No. 6047703
; GENERAL INFORMATION:
;    APPLICANT:  Kinders, Robert J.
;    APPLICANT:  Enfield, David L.
;    APPLICANT:  Hass, G. Michael
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SCREENING
;    TITLE OF INVENTION:  FOR OR MODULATING A TUMOR ASSOCIATED ANTIGEN
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/824,692
;       FILING DATE:  08 APR 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sharkey, Richard G.
;       REGISTRATION NUMBER:  32,629
;       REFERENCE/DOCKET NUMBER:  130001.404
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 622-4900
;       TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO:  42:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  196 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:
;       TOPOLOGY:  linear
US-08-824-692-42

  Query Match              19.6%;  Score 147;  DB 3;  Length 196;
  Best Local Similarity    27.9%;  Pred. No. 1.4e-08;
  Matches   36;  Conservative   22;  Mismatches   61;  Indels   10;  Gaps    5;

Qy        3 SCGSPPPTLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
            || :|: | :| | |          | :|| |   : :  |:: ::|:      :| | :| |
Db       96 SCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCL   - NGNWTEP-P 109

Qy       62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG- 120
            :|:          | :|| |        : |    || : |:  : : |:| : |:  |  |
Db      110 QCKDSTGKCGPPPPTIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NGQWSE 168

Qy      121  PTRLPTCV 128
             |  |  ||
Db      169 PPKCLHPCV 177


RESULT 38
US-08-824-692-41
; Sequence 41, Application US/08824692
; Patent No. 6017703
;  GENERAL INFORMATION:
;    APPLICANT:  Kinders, Robert J.
;    APPLICANT:  Enfield, David L.
;    APPLICANT:  Hass, G. Michael
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SCREENING
;    TITLE OF INVENTION:  FOR OR MODULATING A TUMOR ASSOCIATED ANTIGEN
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/824,692
;      FILING DATE:  08-APR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sharkey, Richard G.
;      REGISTRATION NUMBER:  32,629
;      REFERENCE/DOCKET NUMBER:  130001.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  229 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US-08-824-692-41

  Query Match              19.6%;  Score 147;  DB 3;  Length 229;
  Best Local Similarity    27.9%;  Pred. No. 1.6e-08;
  Matches   36;  Conservative   22;  Mismatches   61;  Indels   10;  Gaps    5;

Qy        3 SCGSPPPTLNGRI-SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLC-TKDKVDGTWDKPAP 61
            || :|: | :| | |    :    | :|| |   : :  |:: :::      :| | :| |
Db      144 SCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCL   - NGNWTEP-P 197

Qy       62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG- 120
            :|:          | :|| |        : |    || : |:  : : |:| : |:  |  |
Db      198 QCKDSTGKCGPPPPTIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NGQWSE 256

Qy      121  PTRLPTCV 128
             |  |  ||
Db      257 PPKCLHPCV 265


RESULT 39
US-08-824-692-29
; Sequence 29, Application US/08824692
; Patent No. 6017703
;  GENERAL INFORMATION:
;    APPLICANT:  Kinders, Robert J.
;    APPLICANT:  Enfield, David L.
;    APPLICANT:  Hass, G. Michael
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR SCREENING
;    TITLE OF INVENTION:  FOR OR MODULATING A TUMOR ASSOCIATED ANTIGEN
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/824,692
;      FILING DATE:  08-APR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sharkey, Richard G.
;      REGISTRATION NUMBER:  32,629
;      REFERENCE/DOCKET NUMBER:  130001.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6041
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  290 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
US 08-824-692-29

  Query Match              19.6%;  Score 147;  DB 3;  Length 290;
  Best Local Similarity    27.9%;  Pred. No. 2.1e-08;
  Matches   36;  Conservative   22;  Mismatches   61;  Indels   10;  Gaps    5;

Qy        3 SCGSPPPTLNGRI-SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLC-TKDKVDGTWDKPAP 61
            || :|: | :| | |    :    | :|| |   : :  |:: :::      :| | :| |
Db      144 SCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCL   - NGNWTEP-P 197

Qy       62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG- 120
            :|:          | :|| |        : |    || : |:  : : |:| : |:  |  |
Db      198 QCKDSTGKCGPPPPTIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NGQWSE 256

Qy      121  PTRLPTCV 128
             |  |  ||
Db      257 PPKCLHPCV 265
```

```
           :|:        | ||  |       : |    || : |:  : : ||| : |: |  |
Db     198 QCKDSTGKCGPPPPITNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NGQWSE 256

Qy     121  PTRLPTCV 128
            |  |  ||
Db     257 PPKCLHPCV 265


RESULT 40
5514582-43
;Patent No. 5514582
;     APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
;     TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;     NUMBER OF SEQUENCES: 43
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/185,670
;       FILING DATE: 21 JAN 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 986,931
;       FILING DATE: 08 DEC-1992
;       APPLICATION NUMBER: 808,122
;       FILING DATE: 16-DEC-1991
;       APPLICATION NUMBER: 440,625
;       FILING DATE: 22 NOV 1989
;       APPLICATION NUMBER: 315,015
;       FILING DATE: 23-FEB 1989
;SEQ ID NO:43:
;       LENGTH: 126
5514582-43

  Query Match             19.2%;  Score 144;  DB 6;  Length 126;
  Best Local Similarity   32.4%;  Pred. No. 1.5e-08;
  Matches   41;  Conservative   13;  Mismatches   59;  Indels   14;  Gaps    5;

Qy       4 CGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPKC 63
           | ||  | ||::      :  |: | |||:  : ||||    |          |: || |
Db       5 CPSPRDIDNGQLDIGG-VEFGSSITYSCNSGYHLIGESKSYCELGSTGSMVWNPEAPIC 62

Qy      64 EYFNKYSSCPFPIVPG---GYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |    |  |  |  |    ||:      |  |  ||::| : :|: ||  | |
Db      63 ESVKCQS--PPSISNGRHNGYE----DFYTDGSVVTYSCNSGYSLIGNSGVLCSGGEWSD 116

Qy     121 PTRLPTC 127
           |   |||
Db     117 P---PTC 120


RESULT 41
US-08-296-014A-4
; Sequence 4, Application US/08296014A
; Patent No. 5716814
;  GENERAL INFORMATION:
;     APPLICANT:  Ding, Jeak Ling
;     APPLICANT:  Ho, Bow
;     TITLE OF INVENTION:  The Cloned Factor C cDNA of the
;     TITLE OF INVENTION:  Singapore Horseshoe Crab, Carcinoscorpius
;     TITLE OF INVENTION:  rotundicauda and Purification of Factor C Proenzyme
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;       STREET:  8110 Gatehouse Road, Suite 500 East
;       CITY:  Falls Church
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22042
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/296,014A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murphy, Jr., Gerald M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1781-105P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1019 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-296-014A-4

  Query Match             19.2%;  Score 144;  DB 1;  Length 1019;
  Best Local Similarity   24.5%;  Pred. No. 2.3e-07;
  Matches   38;  Conservative   31;  Mismatches   50;  Indels   46;  Gaps    8;

Qy       2 ISC-------GSPPPIL----------NGRISYYSTPIAVGTVIRYSCSGIFRLIGEKSL 44
             |||          ||  :           :|::: |  :  |  :|:||   : |||:::|
Db     130 ISCLPNGQWSNFPPKCIRECAMVSSPERGSVNATSGDMIEGATLRFSCDSPYYLIGQEIL 189

Qy      45 LCITKLKVDGTWDKPAPKCEYFNKYSSCP--EELVPGGYKIR-------GSTPYRHGDSV 95
             |    : :| |:   |:|:         ||  :|     :|::       |  |   |
Db     240 TC----QGNGQWNGQIPQCK---NLVFCPDLDEVNHAERKVKIGVEQKYGQFP-QGTEV 290

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  ::  : ' |
Db     291 TYTCSGNYFLMGFDTLKCNPDGSWSGSQ-PSCVEV 324


RESULT 42
US-08-596-405-4
; Sequence 4, Application US/08596405
; Patent No. 5858706
;  GENERAL INFORMATION:
;     APPLICANT:  Ding, Jeak Ling
;     APPLICANT:  Ho, Bow
;     TITLE OF INVENTION:  The Cloned Factor C cDNA of the
;     TITLE OF INVENTION:  Singapore Horseshoe Crab, Carcinoscorpius
;     TITLE OF INVENTION:  rotundicauda and Purification of Factor C Proenzyme
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;       STREET:  8110 Gatehouse Road, Suite 500 East
;       CITY:  Falls Church
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22042
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/596,405
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murphy, Jr., Gerald M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1781-105
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1019 amino acids
```

```
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US 08/596,405 4

  Query Match              19.2%;  Score 144;  DB 2;  Length 1019;
  Best Local Similarity    24.5%;  Pred. No. 2.4e-07;
  Matches   38;  Conservative   41;  Mismatches   50;  Indels   36;  Gaps    8;

Qy       2 ISC      GSPPPIL          NGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSL 44
           |||      ||  :            :|:::  |  :  |  :|:||   : |||:::|
Db     180 ISCLPNGQWSNFPPKCIRECAMVSSPEHGKVNALSGDMIEGATLRFSCDSPYYLIGQETL 239

Qy      45 ICTIKDKVDGIWDKPAPKCEYFNKYSSCP  EPIVPGGYKIR      GSIPYRHGDSV 95
             |    : :| |:   |:|:        ||  :|:    :|::       | |   | |
Db     240 IC    QGNGQWNGQIPQCK   NLVFCPDLDPVNHAEHKVKIGVEQKYGQFP -QGTEV 290

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  :: |:|| |
Db     291 IYICSGNYFLMGFDILKCNPDGSWSGSQ PSCVKV 324


RESULT 43
US 08/877,620 4
; Sequence 4, Application US/08877620
; Patent No. 5985590
;  GENERAL INFORMATION:
;     APPLICANT:  Ding, Jeak Ling
;     APPLICANT:  Ho, Bow
;     TITLE OF INVENTION:  The Cloned Factor C cDNA of the
;     TITLE OF INVENTION:  Singapore Horseshoe Crab, Carcinoscorpius
;     TITLE OF INVENTION:  rotundicauda and Purification of Factor C Proenzyme
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;       STREET:  8110 Gatehouse Road, Suite 500 East
;       CITY:  Falls Church
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22042
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/877,620
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/596,405
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murphy, Jr., Gerald M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1781-105P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1019 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US 08/877,620 4

  Query Match              19.2%;  Score 144;  DB 2;  Length 1019;
  Best Local Similarity    24.5%;  Pred. No. 2.4e-07;
  Matches   38;  Conservative   41;  Mismatches   50;  Indels   36;  Gaps    8;

Qy       2 ISC      GSPPPIL          NGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSL 44
           |||      ||  :            :|:::  |  :  |  :|:||   : |||:::|
Db     180 ISCLPNGQWSNFPPKCIRECAMVSSPEHGKVNALSGDMIEGATLRFSCDSPYYLIGQETL 239

Qy      45 ICTIKDKVDGIWDKPAPKCEYFNKYSSCP  EPIVPGGYKIR      GSIPYRHGDSV 95
             |    : :| |:   |:|:        ||  :|:    :|::       | |   | |
Db     240 IC    QGNGQWNGQIPQCK   NLVFCPDLDPVNHAEHKVKIGVEQKYGQFP -QGTEV 290

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  :: |:|| |
Db     291 IYICSGNYFLMGFDILKCNPDGSWSGSQ PSCVKV 324


RESULT 44
US/08/296,014A/2
; Sequence 2, Application US/08296014A
; Patent No. 5716834
;  GENERAL INFORMATION:
;     APPLICANT:  Ding, Jeak Ling
;     APPLICANT:  Ho, Bow
;     TITLE OF INVENTION:  The Cloned Factor C cDNA of the
;     TITLE OF INVENTION:  Singapore Horseshoe Crab, Carcinoscorpius
;     TITLE OF INVENTION:  rotundicauda and Purification of Factor C Proenzyme
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;       STREET:  8110 Gatehouse Road, Suite 500 East
;       CITY:  Falls Church
;       STATE:  Virginia
;       COUNTRY:  USA
;       ZIP:  22042
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/296,014A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murphy, Jr., Gerald M.
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1781-105P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;       TELEX:  248345
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1083 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US 08-296-014A-2

  Query Match              19.2%;  Score 144;  DB 1;  Length 1083;
  Best Local Similarity    24.5%;  Pred. No. 2.5e-07;
  Matches   38;  Conservative   41;  Mismatches   50;  Indels   36;  Gaps    8;

Qy       2 ISC      GSPPPIL          NGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSL 44
           |||      ||  :            :|:::  |  :  |  :|:||   : |||:::|
Db     244 ISCLPNGQWSNFPPKCIRECAMVSSPEHGKVNALSGDMIEGATLRFSCDSPYYLIGQETL 303

Qy      45 ICTIKDKVDGIWDKPAPKCEYFNKYSSCP  EPIVPGGYKIR      GSIPYRHGDSV 95
             |    : :| |:   |:|:        ||  :|:    :|::       | |   | |
Db     304 IC    QGNGQWNGQIPQCK   NLVFCPDLDPVNHAEHKVKIGVEQKYGQFP -QGTEV 354

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  :: |:|| |
Db     355 IYICSGNYFLMGFDILKCNPDGSWSGSQ PSCVKV 388


RESULT 45
```

```
US 08 596 405 2
; Sequence 2, Application US/08596405
; Patent No. 5858706
;  GENERAL INFORMATION:
;    APPLICANT:  Ding, Jeak Ling
;    APPLICANT:  Ho, Bow
;    TITLE OF INVENTION:  The Cloned Factor C cDNA of the
;    TITLE OF INVENTION:  Singapore Horseshoe Crab, Carcinoscorpius
;    TITLE OF INVENTION:  rotundicauda and Purification of Factor C Proenzyme
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  8110 Gatehouse Road, Suite 500 East
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/596,405
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1781 1050
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 205-8000
;      TELEFAX:  (703) 205-8050
;      TELEX:  248345
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1083 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US 08 596 405 2

  Query Match             19.2%;  Score 144;  DB 2;  Length 1083;
  Best Local Similarity   24.5%;  Pred. No. 2.5e-07;
  Matches   38;  Conservative   41;  Mismatches   50;  Indels   36;  Gaps    8;

Qy       2 ISC---  --GSPPPIL        -- NGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSL 44
             |||       ||  :          :|:::  |  :  |  :|:||   : |||:::|
Db     244 ISCLPNGQWSNFPPKCIRECAMVSSPEHGKVNALSGDMIEGATLRFSCDSPYYLIGQETL 303

Qy      45 ICITKDKVDGTWDKPAPKCEYFNKYSSCP--EPIVPGGYKIR-------GSTPYRHGDSV 95
            |    : :| |:   |:|:       ||  :|:    :|::       |  |   |  |
Db     304 TC  --QGNGQWNGQIPQCK--  NLVFCPDLDPVNHAEHKVKIGVEQKYGQFP--QGTEV 354

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  :: |:|| |
Db     355 TYTCSGNYFLMGPDTLKCNPDGSWSGSQ PSCVKV 388


Search completed: November  6, 2002, 16:06:53
Job time : 14.461 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 26.8996 Seconds
(without alignments)
553.312 Million cell updates/sec

Title: US-09-834-309-4
Perfect score: 751
Sequence: 1 GISCGSPPPILNGRISYYST ANNMWGPTPIPTCVSVFPLE 134

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_042802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	751	100.0	1087	12	AAR11139	B lymphocyte membr
2	219	29.2	1537	12	AAR11982	Partial human comp
3	219	29.2	1929	22	ABP00103	Novel human diagno
4	219	29.2	1930	19	AAW45899	Human complement r
5	219	29.2	2039	12	AAR11810	Human complement 1
6	219	29.2	2039	20	AAY55751	Human C3b/C4b rece
7	219	29.2	2039	22	ABP00287	Novel human diagno
8	219	29.2	2044	22	ABR11782	Human CR1 protein
9	219	29.2	2044	22	AAM39224	Human polypeptide
10	219	29.2	2044	22	AAM41019	Human polypeptide
11	219	29.2	2317	19	AAP92219	CR1 protein. Homo
12	216	28.8	263	22	AAB48847	Cowpox virus Inti
13	215	28.6	263	10	AAP92003	Deduced sequence o
14	215	28.6	263	20	AAY29859	Vaccinia complemen
15	215	28.6	263	21	AAB13074	Complement inhibit
16	212	28.2	2039	14	AAR36743	CR1. Homo sapiens
17	211	28.1	450	20	AAY55755	Human CR1 protein
18	210	28.0	263	20	AAY29858	Vaccinia virus SPI
19	210	28.0	263	20	AAY29860	Mutated VCP giving
20	204	27.2	646	20	AAY55758	Human CR1 protein
21	203.5	27.1	579	19	AAW99924	Amino acid sequenc
22	199	26.5	438	20	AAY55756	Human CR1 protein
23	199	26.5	450	20	AAY55753	Human CR1 protein
24	199	26.5	496	20	AAY55752	Human CR1 protein
25	199	26.5	581	12	AAR13480	Human C4 binding p
26	199	26.5	778	19	AAW73147	Amino acid sequenc
27	198	26.4	133	15	AAR47156	Sequence of solubl
28	198	26.4	254	15	AAR47154	Sequence of solubl
29	198	26.4	450	20	AAY55754	Human CR1 protein
30	198	26.4	458	20	AAY55757	Human CR1 protein
31	196	26.1	481	13	AAR29092	CR1 4 (amino acids
32	196	26.1	483	13	AAR29091	CR1 4 (amino acids
33	196	26.1	543	13	AAR28543	CR1-4 (35E, 47Y) a
34	196	26.1	543	13	AAR28544	CR1-4 (35E) analog
35	196	26.1	543	13	AAR28545	CR1-4 (47Y) analog
36	196	26.1	543	13	AAR28546	CR1 4 (44T, 47D, 4
37	196	26.1	543	13	AAR28547	CR1-4 (52S, 54S, 5
38	196	26.1	543	13	AAR28548	CR1 4 (57V, 59K) a
39	196	26.1	543	13	AAR28549	CR1 4 (64K, 65I) a
40	196	26.1	543	13	AAR28550	CR1-4 (64K) analog
41	196	26.1	543	13	AAR28551	CR1-4 (65T) analog
42	196	26.1	543	13	AAR28552	CR1-4 (78I, 79D) a
43	196	26.1	543	13	AAR28553	CR1-4 (85R, 87N) a
44	196	26.1	543	13	AAR28554	CR1-4 (92I, 94H) a
45	196	26.1	543	13	AAR28555	CR1-4 (92I) analog
46	196	26.1	543	13	AAR28556	CR1-4 (94H) analog
47	196	26.1	543	13	AAR28557	CR1-4 (99H, 103E)
48	196	26.1	543	13	AAR28558	CR1-4 (109N, 110A,
49	196	26.1	543	13	AAR28559	CR1 4 (114 117STKP
50	196	26.1	543	13	AAR28560	CR1-4 (114S) analo
51	196	26.1	543	13	AAR28561	CR1-4 (115I) analo
52	196	26.1	543	13	AAR28562	CR1-4 (116R) analo
53	196	26.1	543	13	AAR28563	CR1-4 (117P) analo
54	196	26.1	543	13	AAR28564	CR1-4 (116R, 117P)
55	196	26.1	543	13	AAR28565	CR1 4 (121g) analo
56	196	26.1	543	13	AAR28566	CR1-4 (418R, 419N)
57	196	26.1	543	13	AAR28567	CR1-4 (418-421 RNP
58	196	26.1	543	13	AAR28568	CR1 4 (447I, 449Y)
59	196	26.1	543	13	AAR28569	CR1-4 (466 474 SIK
60	196	26.1	543	13	AAR28570	CR1-4 (266 274 KIK
61	196	26.1	543	13	AAR28571	CR1-4 (464 467 NAA
62	195	26.0	254	15	AAR47155	Sequence of solubl
63	194	25.8	645	21	AAB52125	Macaca mulatta rha
64	190	25.3	363	18	AAW12414	Porcine complement
65	190	25.3	363	20	AAY30918	MCP protein. Unid
66	172	22.9	1497	22	AAM93954	Human polypeptide,
67	168	22.4	515	22	AAM93955	Human polypeptide,
68	167.5	22.3	497	22	AHB10508	Human cDNA SEQ ID
69	167.5	22.3	497	22	AAU18149	Novel human uterin
70	167.5	22.3	497	22	AAU17348	Human novel secret
71	167.5	22.3	497	22	AAU19960	Novel human calcin
72	167.5	22.3	1139	22	ASB10326	Human cDNA SEQ ID
73	167.5	22.3	1139	22	AAU13126	Novel human uterin
74	167.5	22.3	1139	22	AAU19963	Human novel secret
75	167.5	22.3	1139	22	AAU19902	Novel human calcin
76	166	22.1	357	12	AAR11927	Human membrane cof
77	162	21.6	336	12	AAR15232	CD46 from clone pm
78	162	21.6	350	12	AAR15233	CD46 from clone pm
79	162	21.6	373	12	AAR15230	CD46 from clone pm
80	162	21.6	377	12	AAR15229	CD46 from clone pm
81	159	21.2	254	17	AAW06883	Membrane cofactor
82	159	21.2	314	17	AAR93943	CD46 construct del
83	159	21.2	376	17	AAR93935	CD46 wild type. H
84	159	21.2	377	17	AAR93942	CD46 construct sub

85	159	21.2	377	17	AAR93941	CD46 construct sub
86	159	21.2	377	18	AAW27484	Human MCP. Homo s
87	159	21.2	378	17	AAR93940	CD46 construct sub
88	159	21.2	384	12	AAR10924	Human membrane cof
89	159	21.2	384	16	AAR86316	Human CD46. Homo
90	159	21.2	421	21	AAB58394	Lung cancer associ
91	159	21.2	421	22	AAG75528	Human colon cancer
92	159	21.2	577	17	AAW06882	Membrane co-factor
93	159	21.2	611	22	AAE12569	CAB2 protein. Uni
94	159	21.2	611	22	AAE03762	CAB-2 chimeric pro
95	158.5	21.1	299	17	AAW06881	Decay accelerating
96	158.5	21.1	320	22	AAG68150	Codon modified hum
97	158.5	21.1	376	20	AAY50035	Human complement r
98	158.5	21.1	38	8	AAP70048	Human decay accele
99	158.5	21.1	38	10	AAB94273	Decay accelerating
100	158.5	21.1	38	16	AAR66683	Decay accelerating

ALIGNMENTS

```
RESULT 1
AAR11139
ID   AAR11139 standard; Protein; 1087 AA.
XX
AC   AAR11139;
XX
DT   29-MAY-1991  (first entry)
XX
DE   B-lymphocyte membrane glycoprotein CR2
XX
KW   CR2; B lymphocyte; membrane receptor protein; Epstein-Barr virus;
CC   extracellular domain.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          20..990
FT                   /note= "1"
FT   Region          20..154
FT                   /note= "2"
FT   Region          20..276
FT                   /note= "3"
FT   Region          20..652
FT                   /note= "4"
FT   Region          20..770
FT                   /note= "5"
FT   Region          20..1025
FT                   /note= "6"
FT   Peptide         1..20
FT                   /label= sig peptide
FT   Protein         21..1087
FT                   /label= mat protein
XX
PN   WO9102536-A.
XX
PD   07-MAR-1991.
XX
PF   22-AUG-1990;   90WO-US04817.
XX
PR   23-AUG-1989;   89US-0398224.
XX
PA   (SCRI ) SCRIPPS CLINIC & RES.
XX
PI   Moore MD;  Cooper ND;  Nemerow GR;
XX
DR   WPI; 1991-087109/12.
DR   N-PSDB; AAQ10989.
XX
PT   Synthetic polypeptide(s) of extracellular domain - treat
PT   Epstein Barr virus infectious and diagnose same by formation of
PT   complex between CR2 and EBV
XX
PS   Disclosure; Fig 9.1 9.9; 60pp; English.
XX
CC   The six indicated fragments encode claimed and disclosed
CC   polypeptides, which are synthesised by recombinant expression, prot.
CC   in a baculovirus expression system in which a DNA plasmid, contg. a
CC   cDNA encoding this sequence, is truncated to encode a CR2 polypeptide
CC   that comprises a region of the extracellular domain of CR2.  The
CC   construct is then inserted downstream of an appropriate promoter in
CC   a transfer vector and integrated into a baculovirus which is then
CC   used to infect host insect cells for expression.
CC   The polypeptides correspond to B lymphocyte membrane receptor
CC   protein for Epstein Barr Virus (EBV) and give specific binding.
CC   Protein number 3 is used to inhibit infection of mammalian cells in
CC   contact with an aq. medium, (esp. mammalian blood) and of human
CC   B-lymphocytes, by EBV.  The peptides are used to detect the presence
CC   of EBV in an aq. sample and to detect antibodies directed against CR2.
CC   The peptides are used in a pharmacological compsn. as active
CC   ingredient with a carrier to treat immune disorders.
XX
SQ   Sequence   1087 AA;

  Query Match             100.0%;  Score 791;  DB 12;  Length 1087;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-67;
  Matches  134;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GISCGSPPPTINGPISYYSIPIAVGTVIIVYSCSGIFMICKKRVKKTPWRRPA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 GISCGSPPPTINGPISYYSIPIAVGTVIIVYSCSGIFMICKKRVKKTPWRRPA 79

Qy      61 PRCEYFNKYSSCPEPIVPGGYRIPGSTPYRPEDSVIFACEINFSMNGNKSVWCQANNMWG 120
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 PRCEYINKYSSCPEPIVPGGYRIPGSTPYRPEDSVIFACEINFSMNGNKSVWCQANNMWG 139

Qy     121 PTRLPICVSVFPLE 134
           ||||||||||||||
Db     140 PTRLPTCVSVFPLE 153


RESULT 2
AAR11982
ID   AAR11982 standard; Protein; 1537 AA.
XX
AC   AAR11982;
XX
DT   25-JUN-1991  (first entry)
XX
DE   Partial human complement type 1 receptor.
XX
KW   complement system; C3b/C4b receptor; CR1; allergic reaction;
KW   immune response; long homologous repeat; LHR.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..438
FT                   /label= LHR B
FT   Region          439..891
FT                   /label= LHR C
FT   Region          892..1341
FT                   /label= LHR D
FT   Region          1495..1498
FT                   /note= "positively-charged; preceded by hydrophobic
FT                   sequence"
FT   Region          1521..1526
FT                   /note= "has 67 per cent homology to site of protein
FT                   kinase C phosphorylation in the EGF
FT                   receptor"
XX
PN   WO9105047-A.
XX
PD   18-APR-1991.
XX
PF   25-SEP-1990;   90WO-US05454.
```

```
XX
PR   26-SEP-1989;   89US-041274S.
PR   26-SEP-1990;   90US-091234A.
XX
PA   (TCEL ) T CELL SCI INC.
PA   (UYJO ) JOHNS HOPKINS UNIVERSITY.
PA   (BRIG ) BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI   Fearon DT,  Klickstein LB,  Wong WW,  Carson GR,  Hoh M,  Concino MF;
PI   Makrides SC,  Marsh HC;
XX
DR   WPI; 1991-132854/18.
DR   N-PSDB; AAQ11643.
XX
PT   Human complement receptor type 1 gene, encoded proteins and
PT   fragments - for treatment of immune disorders, myocardial infarct,
PT   damage due to inflammation and in treatment of thrombosis
XX
PS   Disclosure; Fig 5; 234pp; English.
XX
CC   This sequence comprises three of the four tandem, direct, long
CC   homologous repeats of the full-length F allozyme of CR1. LHR-A is
CC   absent. Each LHR might represent a single C3b/C4b binding domain,
CC   making the receptor multivalent. The LHR's are composed of 7 short
CC   consensus repeats of 60-70 residues resembling the SCR's of other
CC   C3/C4 binding proteins. The protein and fragments of it having C3b
CC   and/or C4b binding activity can be used to treat immune disorders
CC   or disorders involving inappropriate complement activity.
CC   See also AAQ11642.
XX
SQ   Sequence   1547 AA;

  Query Match             29.2%;  Score 219;  DB 12;  Length 1547;
  Best Local Similarity   36.7%;  Pred. No. 3.7e-13;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |   |||: | |         | |:|||:| | :||
Db    1015 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy      54 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1073 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1130 QCQTNGRWGP-KLPHCSRV 1147


RESULT 3
ABG00103
ID   ABG00103 standard; Protein; 1929 AA.
XX
AC   ABG00103;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #94.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS64290.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID NO 30462; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30477 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1929 AA;

  Query Match             29.2%;  Score 219;  DB 22;  Length 1929;
  Best Local Similarity   36.7%;  Pred. No. 4.8e-13;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |   |||: | |         | |:|||:| | :||
Db    1407 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1464

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1465 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1521

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1522 QCQTNGRWGP-KLPHCSRV 1539


RESULT 4
AAW45899
ID   AAW45899 standard; peptide; 1930 AA
XX
AC   AAW45899;
XX
DT   30-JUN-1998  (first entry)
XX
DE   Human complement receptor 1 (residues 1-1929).
XX
KW   Membrane binding element; thrombotic disease; soluble protein;
KW   complement-related disease; integral membrane protein; inflammation;
KW   short consensus repeat; SCR 1-3; CR1; complement receptor type 1.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Cross-links     1930
```

```
FT                     /note= "Disulphide linked to Cys in peptide given
FT                      in AAW45889"
XX
PN   WO9802454 A2.
XX
PD   22-JAN-1998.
XX
PF   08-JUL-1997;   97WO-EP03715.
XX
PR   15-JUL-1996;   96GB-0014871.
XX
PA   (ADPR ) ADPROTECH PLC.
XX
PI   Dodd I,  Mossakowska DE1,  Smith RAG;
XX
DR   WPI; 1998-110524/10.
XX
PT   Derivatives of soluble polypeptide(s) bonded to low affinity
PT   membrane binding groups - useful for treating complement related and
PT   thrombotic diseases, providing improved localisation at cellular
PT   membranes
XX
PS   Claim 22; Pages 60-61; 75pp; English.
XX
CC   This sequence represents human complement receptor 1 (CR1, CD 35)
CC   N-terminal fragment. The invention relates to a soluble derivative (A)
CC   of a soluble polypeptide (I), which comprises at least 2 heterologous
CC   membrane binding elements (MBE) of low membrane affinity covalently
CC   associated with (I). MBE interact, independently and with thermodynamic
CC   additivity, with components of cellular or artificial membranes exposed
CC   to extracellular fluids. (A) are used to treat disorders treatable with
CC   (I) itself, specifically inflammation or any other complement-related
CC   disorder (e.g. neurological disease, graft rejection, myocardial
CC   infarction, sepsis, rheumatoid arthritis and many others; including
CC   application to indwelling devices) and thrombolytic disease, but also to
CC   treat allergy, induce weight loss, to treat ischaemia or asthma and as
CC   immune modulators for treating multiple sclerosis. (A) are administered
CC   orally, topically, by injection or inhalation at 0.01-10 (preferably
CC   0.1-10) mg/kg/day.
XX
SQ   Sequence   1930 AA;

  Query Match             29.2%;  Score 219;  DB 19;  Length 1930;
  Best Local Similarity   36.2%;  Pred. No. 4.8e-14;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCQSPPPTLNGRISYYS   TDIAVGTVTKYSC      SGTFRLIGEKSLICTRKDKV 52
           |||  || | ||    :||   |   |||: | |        | :|||:|: | :||
Db    1476 ISCEPPPTISNG  DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1535

Qy      53 DCTWDGFAFKCEYENKYSSGIEPIVIEGGYKTRQSIP YRHGISAVTFACKINFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :  : | |:  | | | |: :|
Db    1536 VGVWSSPPPRCISINK   CIAPEVENAIRVPGNRSFFSLTEITRFRCQPGFVMVGSHTV 1590

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1591 QCQPNGRWGP-KLPHCSRV 1608


RESULT 5
AAR11810
ID   AAR11810 standard; protein; 2039 AA.
XX
AC   AAR11810;
XX
DT   25-JUN-1991  (first entry)
XX
DE   Human complement type 1 receptor.
XX
KW   complement system; C3B/C4B receptor; CR1; allergic reaction;
KW   immune response; clone lambda fl09.1.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..41
FT                   /label= putative signal peptide
FT   Protein         42..2039
FT                   /label= CR1
XX
PN   WO9105047 A.
XX
PD   18-APR-1991.
XX
PF   25-SEP-1990;   90WO-US05454.
XX
PR   26-SEP-1989;   89US-0412745.
PR   26-SEP-1990;   90US-0912349.
XX
PA   (TCEL ) : CELL SCI INC.
PA   (UYJO ) JOHNS HOPKINS UNIVERSITY.
PA   (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI   Fearon DT,  Klickstein LB,  Wong WW,  Carson GR,  Hoh M,  Concino MF;
PI   Makrides SC,  Marsh HC;
XX
DR   WPI; 1991-132854/18.
DR   N-PSDB; AAQ11642.
XX
PT   Human complement receptor type 1 gene, encoded proteins and
PT   fragments - for treatment of immune disorders, myocardial infarct,
PT   damage due to inflammation and in treatment of thrombosis
XX
PS   Claim 41; Fig 1; 234pp; English.
XX
CC   The invention also covers fragments of this protein which have the
CC   ability to bind C3b and/or C4b, have cofactor I activity or can
CC   inhibit C4 or C5 convertase activity. The full length protein, or
CC   its specified fragments are used to treat patients with immune
CC   disorders or a disorder caused by inappropriate complement
CC   activity. The protein is also used to treat thrombotic conditions
CC   in humans and animals. See also AAQ11643.
XX
SQ   Sequence   2039 AA;

  Query Match             29.2%;  Score 219;  DB 12;  Length 2039;
  Best Local Similarity   36.7%;  Pred. No. 5.2e-14;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCQSPPPTLNGRISYYS   TDIAVGTVTKYSC      SGTFRLIGEKSLICTRKDKV 52
           |||  || | ||    :||   |   |||: | |        | :|||:|: | :||
Db    1517 ISCEPPPTISNG -DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 DCTWDGFAFKCEYENKYSSGIEPIVIEGGYKTRQSIP YRHGISAVTFACKINFSMNGNKSV 111
           | |  | |:|   ||   |  | |    :: |:   :  : | |:  | | | |: :|
Db    1575 VGVWSSPPPRCISINK-- CIAPEVENAIRVPGNRSFFSLTEITRFRCQPGFVMVGSHTV 1631

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1632 QCQPNGRWGP-KLPHCSRV 1649


RESULT 6
AAY55751
ID   AAY55751 standard; protein; 2039 AA.
XX
AC   AAY55751;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human C3b/C4b receptor (CR1) protein.
XX
KW   C3b/C4b receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; diagnostic.
```

XX
OS Homo sapiens.
XX
PN US5981481 A.
XX
PD 09 NOV 1999.
XX
PF 06 JUN 1995; 95US-0470652.
XX
PR 03 APR 1989; 89US-0332865.
PR 06 DEC 1974; 74US-0850238.
PR 24 FEB 1993; 93US-0026134.
PR 01 APR 1988; 88US-0176532.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (AVAN) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI Concino MF, Wong WW, Makrides SC, Klickstein LB, Fearon DF, Ip SH;
PI Marsh HC, Carson GR;
XX
DR WPI; 1999-633857/54.
DR N-PSDB; AAZ38150.
XX
PT A human C3B/C4B receptor (CR1) protein having antiinflammatory and
PT cardiant activity
XX
PS Disclosure; Fig 1A-P; 87pp; English.
XX
CC The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC protein or fragment is expressed as a cell-surface protein on the surface
CC of a non-human cell and exhibits a complement regulatory activity of full
CC length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC may be mediated through the inhibition of complement pathway enzymes. The
CC soluble CR1 protein exhibits a complement regulatory activity, and this
CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC and inflammation. The CR1 protein and its fragments can also be used in
CC the treatment of conditions which involve unwanted complement activity,
CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC CR1 antibodies are used in assays, and diagnostics. The present sequence
CC represents the human CR1 protein.
XX
SQ Sequence 2039 AA;

Query Match 29.2%; Score 219; DB 20; Length 2039;
Best Local Similarity 36.7%; Pred. No. 5.2e-13;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCGSPPPILNGRISYYS--TPIAVGTVIRYSC----SGTFRLIGEKSLLCITKDKV 52
Db 1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy 53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
Db 1575 VGVWSSPPPRCISINK--CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy 112 WCQANNMWGPTRLPTCVSV 130
Db 1632 QCQTNGRWGP KLPHCSRV 1649

RESULT 7
ABG00287
ID ABG00287 standard; Protein; 2039 AA.
XX
AC ABG00287;
XX
DT 14-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #278.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS64474
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 30646; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant product on of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating ant bodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30477 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2039 AA;

Query Match 29.2%; Score 219; DB 22; Length 2039;
Best Local Similarity 36.7%; Pred. No. 5.2e-13;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
Db 1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy 53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
Db 1575 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy 112 WCQANNMWGPTRLPTCVSV 130
Db 1632 QCQTNGRWGP KLPHCSRV 1649

RESULT 8
ABB11782
ID ABB11782 standard; peptide; 2044 AA
XX

```
AC   ABB10981;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human CKI protein homologue, SEQ ID NO:2152.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   CNS stimulant; osteopathic; cardiotonic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   07-FEB-2001; 2001WO-US003807.
XX
PR   08-FEB-2000; 2000US-0499914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Tang YT, Li C, Asundi V, Chen R, Ma Y, Qian XP, Ren P, Wang D;
PI   Wang J, Wang Z, Wehrman L, Xu C, Xue AJ, Yang Y, Zhang ;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RD;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; ABA09626.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer.
XX
PS   Claim 20; Page 245-246; 2065pp; English.
XX
CC   Sequences ABB10981 ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225 ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in angiogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention.
XX
SQ   Sequence  2044 AA;

  Query Match             29.2%;  Score 219;  DB 22;  Length 2044
  Best Local Similarity   36.7%;  Pred. No. 5.2e-13;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 IS[S]SFFI[N]RISYYS  IPIAV[S]IVIEY[S]  -  S[H]I[L]I[S]E[S][S]I[S]EIEV 52
           |||  || | ||   :||   |    ||| : |          | |:|:|:| |  ||
Db    1522 ISCLAFFIRNG  DRYSNNKLSRHNGIVVIYQCNIGPDGRQLRLIVGERSAYCISKDDQ 1579

Qy      53 IPTW[R]VAPP[E]YENKYLS[F]EI[V]P[P]YKIS[S]  YPPP[S]CVLEA[P]KLNE[M]N[N]KSV 111
           | |  : :||   ||   | | |    :: |:   :   : : | |:   , | |: :|
Db    1580 VGVWSSPPPRCISTNK-  CIAPEVENAIRVPGNRSFFSLITEIIEFRCQPGFVMVGSHTV 1636

Qy     112 WCQANNMWCPIRLPTCVSV 130
           || |  ||| :|| |  |
Db    1637 QCQINGRWGP-KLPHCSRV 1654


RESULT 9
AAM39224
ID   AAM39224 standard; Protein; 2044 AA.
XX
AC   AAM39224;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2369.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US034263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0633450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Tang YT, Li C, Asundi V, Chen R, Ma Y, Qian XP, Ren P, Wang D;
PI   Wang J, Wang Z, Wehrman L, Xu C, Xue AJ, Yang Y, Zhang ;
PI   Zhao QA, Zhou P, Goodrich R, Drmanac RD;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58480.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries
```

```
XX
PS   Example 4; SEQ ID NO 2369; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AA157798-AA161369) and
CC   the encoded polypeptides (AAM38642 AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   2044 AA;

  Query Match             29.2%;  Score 219;  DB 22;  Length 2044;
  Best Local Similarity   36.7%;  Pred. No. 5.2e-13;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS--TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           ||| || | ||   :||   |   ||| : |         | :|:|:|: | :||
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
            | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1637 QCQTNGRWGP-KLPHCSRV 1654


RESULT 10
AAM41010
ID   AAM41010 standard; Protein; 2044 AA.
XX
AC   AAM41010;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5941.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552417.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620412.
PR   03-AUG-2000; 2000US-0633450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Chen R, Mu Y, Qian XB, Ren F, Wang D;
PI   Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI   Zhao QA, Zhou P, Goodrich R, Ermanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60166.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5941; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AA157798-AA161369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification
XX
SQ   Sequence   2044 AA;

  Query Match             29.2%;  Score 219;  DB 22;  Length 2044;
  Best Local Similarity   36.7%;  Pred. No. 5.2e-13;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------ SGTFRLIGEKSLLCITKDKV 52
           ||| ||  ||   :||   |   ||| : |           | :|:|:|: | :||
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCISKDDQ 1579

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
            | |  | |:|   ||   |  | |    :: |:   :   : : | |:  | | |: :|
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1637 QCQTNGRWGP-KLPHCSRV 1654


RESULT 11
AAP92219
ID   AAP92219 standard; protein; 2317 AA.
XX
AC   AAP92219;
XX
DT   22-FEB-1996  (first entry)
XX
DE   CR1 protein.
XX
KW   Complement  cofactor.
XX
OS   Homo sapiens (human).
XX
FH   Key             Location/Qualifiers
FT   Peptide         10..50
FT                   /label= signal_peptide
XX
PN   WO8909220-A.
XX
PD   05-OCT-1989.
XX
PF   31-MAR-1989;   89WO-US01358.
```

XX
PR 07-APR-1998; 98US-017654.
XX
PA (PCEL-) T CELL SCIENCES INC.
PA (JHUY-) THE JOHNS HOPKINS UNIVERSITY.
PA (BRIG-) THE BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;
XX
DR WPI; 1999-369408/31.
DR N-PSDB; AAN91477.
XX
PT New nucleic acid sequences encoding new CR1 protein and its fragment,
PT for diagnosis and control of complement related immune defects,
PT inflammation, myocardial infarct, etc
XX
PS Claim 1; Fig 1; 101pp; English.
XX
CC This is full length CR1 protein, and shortened forms are new, lacking
CC the transmembrane region. The proteins and fragments bind C3b and/or
CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.
CC In the sequence, x=untranslated region. This has 7 short consensus
CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
CC patent). Deletion mutants, eg LHR A for binding of C4b, and LHR C
CC for C3b binding, be made, and LHR B and C can detect CR2 sequences
CC They are useful in diagnosing and treating immune disorders, and prevent
CC perfusion injury.
XX
SQ Sequence 2317 AA;

Query Match 29.2%; Score 219; DB 10; Length 2317;
Best Local Similarity 36.7%; Pred. No. 6e-13;
Matches 51; Conservative 17; Mismatches 55; Indels 16; Gaps 6;

Qy 2 ISCGSPPPILNGRISYYS TPIAVGTVIKYSC SGTFRLIGEPSIICTIKDKV 52
Db 1526 ISCEPPEILNGQ DFYSNNRTFHNGTVVIYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1583

Qy 53 [illegible] 111
Db 1584 VGVWSGPPPRCISINK CTAPEVENAIRVPGNRSFFSLTEITRFRCQPCFVMVGSHTV 1640

Qy 112 WCQANNMWGPTRLPTCVSV 130
Db 1641 QCQTNGRWGP KIPHCSDV 1658

RESULT 12
AAB40846
ID AAB40846 standard; Protein; 263 AA.
XX
AC AAB40846;
XX
DT 13-MAR-2001 (first entry)
XX
DE Cowpox virus inflammation modulatory protein (IMP).
XX
KW Inflammation modulatory protein; IMP; cowpox virus;
KW viral complement control protein; DAF homologue;
KW complement activation inhibition; xenograft rejection.
XX
OS Cowpox virus.
XX
PN WO200071147-A1
XX
PD 30-NOV-2000.
XX
PF 24-MAY-2000; 2000WO-US14203.
XX
PR 25-MAY-1999; 99US-0136134.
XX
PA (KING-) KING FAISAL SPECIALIST HOSPITAL & RES CE.
PA (UYVI-) UNIV VIRGINIA PATENT FOUND INC.
XX
PI Kotwal GJ, Al-Mohanna F, Parhar R;
XX
DR WPI; 2001-349865/36.
DR N-PSDB; AAV87530.
XX
PT Use of viral inflammation modulatory protein for blocking xenograft
PT rejection in a patient
XX
PS Claim 1; Page 26; 71pp; English.
XX
CC The invention relates to a novel method of preventing or treating
CC xenograft rejection in a patient. The method comprises administering an
CC inflammation modulatory protein (IMP) from cowpox virus, which acts as
CC an inhibitor of complement activation. The invention also relates to
CC a pharmaceutical composition comprising the IMP, suitable for the
CC prevention or treatment of xenograft rejection. IMP is structurally
CC related to the family of complement control proteins that includes DAF,
CC and is functionally similar to the soluble complement receptor 1 (sCR1).
CC IMP can block complement activation at a very early stage by binding to
CC C3b and C4b, by inhibiting C3 convertase formation by either the
CC classical or alternative pathway, and by accelerating the decay of the
CC C3 convertase. Unlike other complement inhibitors, IMP retains most of
CC the essential domains for complement control due to evolutionary
CC pressure, and unlike sCR1, cleaves C3 at only the first cleavage site in
CC the presence of factor 1. In addition, the viral protein is very small
CC and is not recognised as a foreign protein because of its close
CC structural similarity to its human homologues. The method and composition
CC of the invention are useful for treating and preventing xenograft
CC rejection in a patient in need of such treatment. The present sequence
CC represents cowpox virus IMP.
XX
SQ Sequence 263 AA;

Query Match 28.9%; Score 216; DB 22; Length 263;
Best Local Similarity 34.4%; Pred. No. 8.4e-14;
Matches 44; Conservative 17; Mismatches 55; Indels 12; Gaps 5;

Qy 2 ISCGSPPPILNGRISYYSTPIAVGTVIKYSCSGTFRLIGEPSIICTIKDKVFWDPEAP 61
Db 146 VKCPSPPSVINGKHNGYEGFYIDGSVVIYSCNSGYSLIGNSTIVC SGGEWSDPLP 199

Qy 62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDTVIFACKTNFSMNGNKSVWCQANNMWG 120
Db 200 TCQIV FCPHPTISNGYLSSTFYRVGSHNCNVTCFEG YRLTGSVSSTCQSNGTWS 255

Qy 121 PTRLPTCV 128
Db 256 PVELPKCV 262

RESULT 13
AAP92003
ID AAP92003 standard; protein; 263 AA.
XX
AC AAP92003;
XX
DT 17-DEC-2001 (updated)
DT 09-FEB-1990 (first entry)
XX
DE Deduced sequence of complement 4b (C4b) binding protein
XX
KW Vaccinia virus WR strain; anti-complement protein; complement
KW inactivators; complement 4b; C4b
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= Signal_peptide
FT Duplication 20
FT /note= "See note a in comments below."
FT Duplication 83
FT /note= "See note a in comments below."
FT Duplication 144

```
FT                   /note= "See note a in comments below."
FT   Duplication     202
FT                   /note= "See note a in comments below."
XX
PN   US5739208-N.
XX
PD   14-MAR-1989.
XX
PF   20-AUG-1988;   88US-0239208.
XX
PR   20-AUG-1988;   88US-0239208.
XX
PA   (USSH ) NAT INST OF HEALTH.
XX
PI   Korwal G;
XX
DR   WPI; 1989-165451/22.
DR   N-PSDB; AAN90113.
XX
PT   New protein with anti-complement activity
PT     encoded by Vaccinia virus 35K gene
XX
PS   Disclosure; Figure 2A; 20pp; English.
XX
CC   C4b binding protein which specifically blocks human complement cascades.
CC   It is the deduced sequence of a 35kDa protein encoded by sequence 52-840
CC   of the 35K gene of vaccinia virus strain WR. Note a - these sites
CC   indicate the start of 60 amino acid tandem repeating units which have a
CC   consensus sequence. The signal peptide sequence is not found in purified
CC   35K protein recovered from the medium of cells infected with vaccinia
CC   virus strain WR. A suggested use is to treat diseases due to abnormally
CC   high complement activity.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ   Sequence   263 AA;

  Query Match             28.6%;  Score 215;  DB 10;  Length 263;
  Best Local Similarity   45.9%;  Pred. No. 1.1e-13;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : | ||| | ||| : |      |:|: ||| :  : |||   :|| |  | |  | |
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLC---  -SGGEWSDP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   |    | : |:| | ||  : :::| |  |   | |
Db     200 TCQIV    -KCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P ELPKCV 262


RESULT 14
AAY29859
ID   AAY29859 standard; Protein; 263 AA.
XX
AC   AAY29859;
XX
DT   16-NOV-1999  (first entry)
XX
DE   Vaccinia complement control protein sequence.
XX
KW   Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW   fusion protein; hyperacute rejection; xenograft; inflammation;
KW   post ischaemic reperfusion injury; malignancy; autoimmune disease;
KW   immune system disorder; neurodegeneration; infection; gene therapy;
KW   blood additive; extracorporeal circulation system.
XX
OS   Vaccinia virus.
XX
OS   Synthetic.
XX
PN   WO9944625-A1.
XX
PD   10-SEP-1999.
XX
PF   02-MAR-1999;   99WO-US04635.
XX
PR   03-MAR-1998;   98US-0076821.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Rosengard AM,  Ahearn JM;
XX
DR   WPI; 1999-550981/46.
DR   N-PSDB; AAZ21091.
XX
PT   New smallpox inhibitor of complement enzyme protein, used to treat
PT   complement-mediated disease, particularly hyperacute rejection
XX
PS   Claim 15; Fig 1; 88pp; English.
XX
CC   The present invention describes the Vaccinia virus smallpox inhibitor of
CC   complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC   activation, and so can be used to treat or prevent complement-mediated
CC   disorders, especially hyperacute rejection, inflammation or post
CC   ischaemic reperfusion injury, malignancies, autoimmune diseases,
CC   immune system disorders, neurodegeneration and infections. Hyperacute
CC   rejection may also be prevented by treating the graft with SPICE before
CC   transplanting it or by using a xenograft that has been transformed to
CC   express SPICE from a gene therapy vector. SPICE is also useful as
CC   additive to blood, e.g. in an extracorporeal circulation system (coated
CC   on tubing) or in storage, also for studying complement activation.
CC   Transgenic animals that express SPICE are used as sources of xenografts.
CC   The present sequence represents a vaccinia complement control protein
CC   (VCP) encoded by the specifically claimed mutated VCP nucleotide
CC   sequence, having a silent T to A transversion at nucleotide position
CC   number 267.
XX
SQ   Sequence   263 AA;

  Query Match             28.6%;  Score 215;  DB 20;  Length 263;
  Best Local Similarity   45.9%;  Pred. No. 1.1e-13;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGIFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : | ||| | ||| : |      |:|: ||| :  : |||   :|| |  | |  | |
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSEYSLIGNSGVLC-----SGGEWSDP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   |    | : |:| | ||  : :::| |  |   | |
Db     200 TCQIV----KCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 15
AAB13014
ID   AAB13014 standard; protein; 263 AA.
XX
AC   AAB13014;
XX
DT   11-DEC-2000  (first entry)
XX
DE   Complement inhibitory protein VCP amino acid sequence.
XX
KW   Alzheimer's disease; Vaccinia virus  VCP; complement pathway inhibitor;
KW   treatment; diagnosis; amyloid plaque.
XX
OS   Vaccinia virus.
```

```
XX
PN   WO200043027-A1.
XX
PD   27-JUL-2000.
XX
PF   19-JAN-2000; 2000WO-US01115.
XX
PR   19-JAN-1999;  99US-0116428.
XX
PA   (UYLO-) UNIV LOUISVILLE RES FOUND INC.
PA   (KOTW/) KOTWAL G J.
PA   (DAIY/) DAILY J.
XX
PI   Kotwal GJ, Daily J;
XX
DR   WPI; 2000-470187/41.
DR   N-PSDB; AAA72748.
XX
PT   Treating Alzheimer's disease using a Vaccinia virus protein that blocks
PT   the complementation pathway
XX
PS   Claim 1; Page 42-43; 96pp; English.
XX
CC   This invention relates to a method for treating Alzheimer's disease. The
CC   method uses a composition comprising a Vaccinia virus complement control
CC   protein, which blocks the complement pathway by binding to complement
CC   components. The protein designated VCP, blocks complement activation and
CC   can bind to amyloid plaques in samples. The transmembrane domain and a
CC   portion of the C terminus (Abeta) of the amyloid precursor protein (APP)
CC   form the nucleus of the amyloid plaque, the hallmark of Alzheimer's
CC   disease. The Abeta peptide activates the complement pathway. The VCP
CC   protein used in the method of the invention, down regulates the
CC   complement pathway activation caused by Abeta. The method and composition
CC   may be used for the treatment of Alzheimer's disease. The VCP protein
CC   also binds to amyloid plaques and may be labelled and used to detect the
CC   presence of amyloid plaques in pathological samples, and therefore
CC   diagnose Alzheimer's disease. The present sequence represents the
CC   Vaccinia virus VCP protein used in the method of the invention.
XX
SQ   Sequence  263 AA;

  Query Match             28.9%;  Score 215;  DB 21;  Length 263;
  Best Local Similarity   35.4%;  Pred. No. 1.1e-13;
  Matches   46;  Conservative   15;  Mismatches   56;  Indels   12;  Gaps    5;

Qy       2 ISCGSEPPEIPNGRISYYSIETIAVGTVIKYSC SGTERELIGEKSLICITKDKVDGTWDKPAP 61
           | | | | ||| | |       ||| ||||  | |||   :||       | |  | |
Db     146 VKCQSLEPSTSNGKHNGYTIDKYTIDGSVVTYSCNSGYSLIGNSGVIC   SGDPWSDP P 199

Qy      62 KCEYFNKYSSCPPPTIPDGYKTRC STPYYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           ||      |  |  |  |    | | |:|| | ||  : |:|| |  |  | |
Db     200 ICQIV    KCPHPITSNGVISSGPKRSYSYNDNVIFKCKYGYKISGSSSSICSPDNIWK 256

Qy     121 PTRLPTCV 128
           |  | |
Db     257 PELPKCV 262


RESULT 10
AAK06743
ID   AAK06743 standard; Protein; 2039 AA.
XX
AC   AAK06743;
XX
DT   14-SEP-1998  (first entry)
XX
DE   CR1.
XX
KW   C3b/C4b receptor; CR1; erythrocyte; monocyte; macrophage; granulocyte;
KW   iC3b; T cell; splenic follicular dendritic cell; soluble; complement;
KW   glomerular podocyte; B cell; C3b; C4b; inactivated C3b; phagocytosis;
KW   plasma; ligand binding activity; immune complex; activator; allotype;
KW   endocytosis; lymphocyte; classical; alternative; pathway; cofactor; F;
KW   C3/C5 convertase; factor I; cleavage; factor I; regulation; glycoprotein;
KW   S; A; B; glycosylation; duplication; repetitive intervening sequence;
KW   endoglycosidase F.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..40
FT                   /note= "Signal peptide"
FT   Protein         503..2039
FT                   /note= "CR1 encoded by CR1 cDNA"
FT   Region          503..990
FT                   /note= "Long homologous repeat A"
FT   Region          491..1393
FT                   /note= "Long homologous repeat B"
FT   Region          1394..1843
FT                   /note= "Long homologous repeat C"
FT   Misc-difference 161..162
FT                   /note= "Unclear in specification"
XX
PN   US5212071-A.
XX
PD   18-MAY-1993.
XX
PF   01-APR-1988;   88US-0176542.
XX
PR   01-APR-1988;   88US-0176542.
PR   03-APR-1989;   89US-0332865.
XX
PA   (BGHM-) BRIGHAM & WOMENS HOSPITAL.
PA   (ICEL-) T CELL SCI INC.
PA   (UYJO-) UNIV JOHNS HOPKINS.
XX
PI   Carson GR, Concino MF, Fearon DT, Ip SH, Klickstein LB;
PI   Makrides SC, Wong WW;
XX
DR   WPI; 1993-175454/21.
DR   N-PSDB; AAQ41867.
XX
PT   Nucleic acid encoding polypeptide having complement regulatory
PT   activity - used to prevent reperfusion injury, inhibit Arthus
PT   reaction and neutrophil mediated tissue damage and reduce
PT   myocardial infarct size and inflammation
XX
PS   Claim 1; Fig 1; 90pp; English.
XX
CC   This sequence represents the entire amino acid sequence of the C3b/C4b
CC   receptor (CR1). CR1 is present on erythrocytes, monocytes/macro
CC   phages, granulocytes, B cells, some T cells, splenic follicular
CC   dendritic cells and glomerular podocytes. CR1 specifically binds
CC   C3b, C4b and inactivated C3b (iC4b). A soluble form of the receptor
CC   is found in plasma which has ligand binding activity and the same
CC   molecular weight as membrane-associated CR1. CR1 binds C3b and C4b
CC   that have covalently attached to immune complexes and other complement
CC   activators. The consequences of these interactions depends on the
CC   type of bearing the receptor. Erythrocyte CR1 binds immune complexes
CC   for transport to the liver. CR1 on neutrophils and monocytes
CC   internalises bound complexes, either by adsorptive endocytosis
CC   or by phagocytotsis. The function of CR1 on B lymphocytes is less
CC   well defined. CR1 can inhibit the classical and alternative pathway
CC   C3/C5 convertases and act as a cofactor for the cleavage of C3b and
CC   C4b by factor I, therefore CR1 has a complement regulatory function
CC   as well as acting as a receptor. CR1 is a glycoprotein composed of
CC   a single polypeptide chain. Four allotypic forms of CR1 have been
CC   identified, differing by increments of approx. 40-50 kD. The two most
CC   common forms, the F and S allotypes, also termed A and B allotypes,
CC   have molecular weights of 250 and 290 kD respectively. The two rarer
CC   forms have molecular weights of 210 and 290 kD. These differences
CC   represent differences in the polypeptide chain of CR1, rather than
CC   glycosylation state because they are not abolished by treatment of
CC   purified receptor portein with endoglycosidase F. The CR1 gene has
CC   been shown to have repetitive intervening sequences which may have
CC   been duplicated in the formation of the larger allotypes.
```

```
XX
SQ   Sequence   2039 AA;

  Query Match             28.2%;  Score 212;  DB 14;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 2.6e-12;
  Matches   50;  Conservative   19;  Mismatches   54;  Indels   16;  Gaps    7;

Qy       2 ISCGSPPPTLNGRISYYS--TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  ||| :: |  :||   |    ||| : | |          |:|:|||:|: | :||
Db    1517 ISC EPPPTISNR DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFQLVGERSIYCTSKDDQ 1574

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   ||   | | |    :: |:   :  : : | |   | | |: :|
Db    1575 VGVWSSPPPRCISTNK--CTAPEVQNAIRVPGNRSFFSLTEIIRFRCHPGFVMVGSHTV 1631

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1632 QCQTNGRWGP KLPHCSRV 1649


RESULT 17
AAY55755
ID   AAY55755 standard; Protein; 450 AA.
XX
AC   AAY55755;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR D SCR fragment.
XX
KW   C4B/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   04-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0530248.
PR   24-FEB-1993;   93US-0026144.
PR   01-APR-1988;   88US-0176542.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN ) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-643857/54.
XX
PT   A human C4B/C4B receptor (CR1) protein having antiinflammatory and
PT   cardiant activity -
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C4B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size.
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR) D sequence.
XX
SQ   Sequence   450 AA;

  Query Match             28.1%;  Score 21.  DB 20;  Length 450;
  Best Local Similarity   36.0%;  Pred. No. 5.2e-13;
  Matches   50;  Conservative   17;  Mismatches   56;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYS---TPIAVGTVIRYSC------ SGTFRLIGEKSLLCITKDKV 52
           |||  | | ||   :||   |    ||| : | |          | :|:|:|: | :||
Db     124 ISCEPPPTISNG--DFYSNNRTSFHNGIVVTYQCHIGPDGEQLFELVGERSIYCISKDDQ 181

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   ||   |    |   :: |:   :  : : | |:  | | |: :|
Db     182 VGVWSSPGPRCISTNK---CTAENVENAIEVPGNRSFFSLTEIIRFRCQPGFVMVGSHIV 238

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db     239 QCQTNGRWGP-KLPHCSRV 256


RESULT 18
AAY29858
ID   AAY29858 standard; Protein; 263 AA.
XX
AC   AAY29858;
XX
DT   16-NOV-1999  (first entry)
XX
DE   Vaccinia virus SPICE protein.
XX
KW   Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW   fusion protein; hyperacute rejection; xenograft; inflammation;
KW   post-ischaemic reperfusion injury; malignancy; autoimmune disease;
KW   immune system disorder; neurodegeneration; infection; gene therapy;
KW   blood additive; extracorporeal circulation system.
XX
OS   Vaccinia virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 13
FT                   /note= "encoded by CGA"
XX
PN   WO9944625-A1.
XX
PD   10-SEP-1999.
XX
PF   02-MAR-1999;   99WO-US04635.
XX
PR   03-MAR-1998;   98US-0076821.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Rosengard AM,  Ahearn JM;
XX
DR   WPI; 1999-550981/46.
DR   N-PSDB; AAZ21090.
XX
PT   New smallpox inhibitor of complement enzyme protein, used to treat
PT   complement-mediated disease, particularly hyperacute rejection
XX
PS   Claim 1; Fig 2; 88pp; English.
XX
CC   The present sequence represents Vaccinia virus smallpox inhibitor of
CC   complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC   activation, and so can be used to treat or prevent complement mediated
CC   disorders, especially hyperacute rejection, inflammation or post
CC   ischaemic reperfusion injury, malignancies, autoimmune diseases.
```

```
CC   immune system disorders, neurodegeneration and infections. Hyperacute
CC   rejection may also be prevented by treating the graft with SPICE before
CC   transplanting it or by using a xenograft that has been transformed to
CC   express SPICE from a gene therapy vector. SPICE is also useful as
CC   additive to blood, e.g. in an extracorporeal circulation system (coated
CC   on tubing) or in storage, also for studying complement activation.
CC   Transgenic animals that express SPICE are used as sources of xenografts.
XX
SQ   Sequence   263 AA;

  Query Match             28.0%;  Score 210;  DB 20;  Length 263;
  Best Local Similarity   45.2%;  Pred. No. 3.4e-13;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 61
           : |  || : |:| : |:     |:|: |||:  : |||   :||       | | | | |
Db     146 VKCQLPPSISNGKHNGYNDFYIDGSVVIYSCNSGYSLIGNSGVLC----  SGGEWSNP P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   :  | : |:| | ||  : :::: | |   | | | ||
Db     200 TCQIV    KCPHPTILNGYISSGEKRSYSYNDNVDFICKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P ELPKCV 262


RESULT 19
AAY29860
ID   AAY29860 standard; Protein; 263 AA.
XX
AC   AAY29860;
XX
DT   10-NOV-1999  (first entry)
XX
DE   Mutated VCP giving SPICE protein sequence.
XX
KW   Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW   fusion protein; hyperacute rejection; xenograft; inflammation;
KW   post ischaemic reperfusion injury; malignancy; autoimmune disease;
KW   immune system disorder; neurodegeneration; infection; gene therapy;
KW   blood additive; extracorporeal circulation system.
XX
OS   Vaccinia virus.
OS   Synthetic.
XX
PN   WO9944625-A1.
XX
PD   10-SEP-1999.
XX
PF   02-MAR-1999;   99WO-US004665.
XX
PR   03-MAR-1998;   98US-0076821.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (UYPI ) UNIV PITTSBURGH.
XX
PI   Rosenstard AM,  Ahearn JM,
XX
DR   WPI; 1999-550981/46.
DR   N-PSDB; AAZ21092.
XX
PT   New smallpox inhibitor of complement enzyme protein, used to treat
PT   complement mediated disease, particularly hyperacute rejection -
XX
PS   Disclosure; Fig 3; 88pp; English.
XX
CC   The present invention describes the Vaccinia virus smallpox inhibitor of
CC   complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC   activation, and so can be used to treat or prevent complement mediated
CC   disorders, especially hyperacute rejection, inflammation or post
CC   ischaemic reperfusion injury, malignancies, autoimmune diseases,
CC   immune system disorders, neurodegeneration and infections. Hyperacute
CC   rejection may also be prevented by treating the graft with SPICE before
CC   transplanting it or by using a xenograft that has been transformed to
CC   express SPICE from a gene therapy vector. SPICE is also useful as
CC   additive to blood, e.g. in an extracorporeal circulation system (coated
CC   on tubing) or in storage, also for studying complement activation.
CC   Transgenic animals that express SPICE are used as sources of xenografts.
CC   The present sequence represents the protein encoded by a specifically
CC   claimed mutated vaccinia complement control protein (VCP) encoding SPICE,
CC   in which position 5 of the protein is occupied by Ser (from VCP) rather
CC   than Arg.
XX
SQ   Sequence   263 AA;

  Query Match             28.0%;  Score 210;  DB 20;  Length 263;
  Best Local Similarity   45.2%;  Pred. No. 3.4e-13;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 61
           : |  || : |:| : |:     |:|: |||:  : |||   :||       | | | | |
Db     146 VKCQLPPSISNGRHNGYNDFYIDGSVVIYSCNSGYSLIGNSGVLC----  SGGEWSNP P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   :  | : |:| | ||  : :::: | |   | | | ||
Db     200 TCQIV    KCPHPTILNGYISSGEKRSYSYNDNVDFICKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P ELPKCV 262


RESULT 20
AAY55758
ID   AAY55758 standard; Protein; 646 AA.
XX
AC   AAY55758;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR-D fragment.
XX
KW   C3B/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury, Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0530238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176532.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-633357/54.
XX
PT   A human C3b/C4b receptor (CR1) protein having antiinflammatory and
PT   cardiant activity
XX
PS   Disclosure; Fig 5B; 87pp; English.
XX
CC   The invention relates to a human C3b/C4b receptor (CR1) protein. The CR1
```

```
CC   protein or fragment is expressed as a cell-surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the human CR1 protein long homologous repeat (LHR)-D fragment.
XX
SQ   Sequence   646 AA;

  Query Match             27.2%;  Score 204;  DB 20;  Length 646;
  Best Local Similarity   35.4%;  Pred. No. 4.1e-12;
  Matches   49;  Conservative   16;  Mismatches   58;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS-- TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           ||| ||   ||   :||   |    ||| | |           | |:|  |: | :||
Db     124 ISCEPPPTSANG--DFYSNNRISFHNGTVVTYQCHTGPDGEQLFELVGRPSIYCTSKDDQ 181

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSI-PYRHGDSVTFACKTNFSMNGNKSV 111
           | | | |:|   ||   | | |    :: |:   :   : : | |:  | | |: :|
Db     182 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRFFFSLTEIIRFRCQPGFVMVGSHTV 238

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db     239 QCQTNGRWGP-KLPHCSRV 256


RESULT 21
AAW89924
ID   AAW89924 standard; Protein; 579 AA.
XX
AC   AAW89924;
XX
DT   19-JUN-1998  (first entry)
XX
DE   Amino acid sequence of a mouse sperm protein designated sp56.
XX
KW   Sperm protein; sp56; protein receptor; Sushi domain;
KW   species restricted epitope; sp56 antigenic determinant;
KW   immunocontraceptive; vaccine; sperm fertility; sp56 antigen.
XX
FH   Key             Location/Qualifiers
FT   Domain          33..91
FT                   /note= "Sushi domain 1"
FT   Domain          96..153
FT                   /note= "Sushi domain 2"
FT   Domain          158..250
FT                   /note= "Sushi domain 3"
FT   Domain          223..278
FT                   /note= "Sushi domain 4"
FT   Domain          283..345
FT                   /note= "Sushi domain 5"
FT   Domain          345..411
FT                   /note= "Sushi domain 6"
FT   Domain          45..509
FT                   /note= "Sushi domain 7"
FT   Domain          412..487
FT                   /note= "sp56 specific sequence domain"
FT   Domain          415..430
FT                   /note= "sp56 specific sequence domain"
FT   Domain          428..443
FT                   /note= "sp56 specific sequence domain"
FT   Domain          438..453
FT                   /note= "sp56 specific sequence domain"
XX
OS   Mus sp.
XX
PN   WO9800440-A1.
XX
PD   08-JAN-1998.
XX
PF   03-JUL-1997;   97WO-US11719.
XX
PR   03-JUL-1996;   96US-0674895.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Bleil JD,  Bookbinder LH,  Hein MB;
XX
DR   WPI; 1998-086898/08.
DR   N-PSDB; AAV09901.
XX
PT   New immunocontraceptive vaccine compositions - containing an
PT   antigenic polypeptide having an sp56 antigenic determinant to induce
PT   anti-sp56 antibodies in mammals
XX
PS   Claim 6; Pages 95-97; 123pp; English.
XX
CC   The present sequence represents a murine sperm protein designated sp56.
CC   The primary sequence of the sp56 protein indicates that it is a
CC   member of a superfamily of protein receptors, characterised by the
CC   presence of a series of 60 amino acid long homology repeats called Sushi
CC   domains. Each Sushi domain contains 10 virtually invariant amino acids.
CC   A stretch of 44 amino acids near the carboxy terminus is believed to
CC   define a species restricted epitope. sp56 antigenic determinants are used
CC   in an immunocontraceptive vaccine composition. The vaccine composition
CC   induces anti-sp56 antibodies in a mammalian subject, in combination with
CC   an excipient. The compositions can be used for oral immunocontraception
CC   for mammals. The method can also be used for assessing sperm fertility,
CC   detection of sperm sp56 antigens and confirming effectiveness of the
CC   vaccines at inducing immune responses. The immunocontraceptive can be
CC   formulated to contain species-restricted sp56 sequences such that
CC   contraception is only achieved in the target species, and not in other
CC   animals or humans who inadvertently consume the bait.
XX
SQ   Sequence   579 AA;

  Query Match             27.1%;  Score 203.5;  DB 19;  Length 579;
  Best Local Similarity   33.1%;  Pred. No. 4.1e-12;
  Matches   42;  Conservative   22;  Mismatches   56;  Indels    7;  Gaps    4;

Qy       3 SCGSPPPILNGRISYYSTPI-AVGTVIRYSCS-TFRLIGEKSLLCITKDKVDGTWDKPAP 61
           :|  |||| || |     :   |:|: |:-  || |:   |::|   ::  | |    |
Db     157 TCEPPPPISNGKHSGRDDDLYTFGSVVIYNCIPTFTLLGNASIVCTVVNKTVGVWRPHPP 216

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSIP-YRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:      |  | :| ||   |      :   ::    ||  | : |:   : |:|| :
Db     217 ACQKI----VCHRPQIPKGYIAPGFRQFYAYEDALEIRCKKGFILRGSSVIRCEANGEWF 272

Qy     121 PTRLPTC 127
           |: :|||
Db     273 PS-IPTC 278


RESULT 22
AAY55756
ID   AAY55756 standard; Protein; 438 AA.
XX
AC   AAY55756;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR-B fragment.
XX
KW   C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR.
XX
OS   Homo sapiens.
```

```
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0530238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176542.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN ) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-633457/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having anti-inflammatory and
PT   cardiant activity
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analoques, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR) B sequence.
XX
SQ   Sequence   450 AA;

  Query Match             26.5%;  Score 199;  DB 20;  Length 450;
  Best Local Similarity   38.0%;  Pred. No. 8.3e-12;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSIPIAVGTVIRYSC  SG  IERLIGEKSLIC*IIKDKVDG 54
           |  || || | ||  ||        |:|: | |  ||   | |:|| |: | : |  |
Db     122 IPCGLPPLIANGDFISTNRENFHYGSVVTYRCNPGSGGKKVFELVGEPSIYCTSNDDQVG 181

Qy      55 IWDKEAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |:|:|   ||   | | | |  :  : : :   : | | |:  | | | : | |
Db     182 IWSGEAPQCIIPNK  CTFPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 228

Qy     114 QANNMWGPTRLPICVSV 130
           || |   :  :|:|  |
Db     229 QALNKWEP ELPSCSRV 244


RESULT 23
AAY55753
ID   AAY55753 standard; Protein; 450 AA.
XX
AC   AAY55753;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR B SCR fragment.
XX
KW   C3B/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0530238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176542.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN ) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-633457/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having anti-inflammatory and
PT   cardiant activity
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analoques, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR) B sequence.
XX
SQ   Sequence   450 AA;

  Query Match             26.5%;  Score 199;  DB 20;  Length 450;
  Best Local Similarity   38.0%;  Pred. No. 8.3e-12;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSIPIAVGTVIRYSC  SG  IERLIGEKSLIC*IIKDKVDG 54
           |  || || | ||  ||        |:|: | |  ||   | |:|| |: | : |  |
Db     122 IPCGLPPLIANGDFISTNRENFHYGSVVTYRCNPGSGGKKVFELVGEPSIYCTSNDDQVG 181

Qy      55 IWDKEAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |:|:|   ||   | | | |  :  : : :   : | | |:  | | | : | |
Db     182 IWSGEAPQCIIPNK  CTFPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 228

Qy     114 QANNMWGPTRLPICVSV 130
           || |   :  :|:|  |
Db     229 QALNKWEP ELPSCSRV 244


RESULT 24
AAY55752
ID   AAY55752 standard; Protein; 496 AA.
XX
AC   AAY55752;
XX
DT   22-FEB-2000  (first entry)
XX
```

```
DE   Human CR1 protein LHR A SCR fragment.
XX
KW   C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0530238.
PR   24-FEB-1994;   94US-0026144.
PR   01-APR-1988;   88US-0176542.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN ) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Pearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-643857/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having antiinflammatory and
PT   cardiant activity -
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell-surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR)-A sequence.
XX
SQ   Sequence   496 AA;

  Query Match             26.5%;  Score 199;  DB 20;  Length 496;
  Best Local Similarity   38.0%;  Pred. No. 9.6e-12;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR ISYYSTPIAVGTVIRYSC---SG---TFRLIGEKSLICITKDKVDC 54
           | || || | ||  ||         |:|: | |   ||    | |:|| |: | : |   |
Db     170 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 229

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | |||:|   ||  | | | |  : : : :   : | | |:  | | | : | |
Db     230 IWSGPAPQCIIPNK-  CTPPNVENGIIVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 286

Qy     114 QANNMWGPTRIPTCVSV 130
           || | | |  |||:|  |
Db     287 QALNKWEP ELPSCSRV 302


RESULT 25
AAR13490
ID   AAR13490 standard; Protein; 581 AA.
XX
AC   AAR13490;
XX
DT   30-OCT-1991  (first entry)
XX
DE   Human C4 binding protein.
XX
KW   C4bp; monomer; complement protein; p-CL.C4bp.1; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..32
FT                   /label= signal_peptide
FT   Protein         33..581
FT                   /label= C4bp
FT   Region          33..93
FT                   /label= SCR8
FT   Region          94..155
FT                   /label= SCR7
FT   Region          156..219
FT                   /label= SCR6
FT   Region          220..279
FT                   /label= SCR5
FT   Region          280..345
FT                   /label= SCR4
FT   Region          346..406
FT                   /label= SCR3
FT   Region          407..464
FT                   /label= SCR2
FT   Region          465..523
FT                   /label= SCR1
FT   Domain          524..581
FT                   /label= C4bp_core
FT                   /note= "responsible for multimer assembly"
FT   Disulfide-bond  34..80
FT                   /note= "intradomain"
FT   Disulfide-bond  65..92
FT                   /note= "intradomain"
XX
PN   WO9111461-A.
XX
PD   08-AUG-1991.
XX
PF   28-JAN-1991;   91WO-US00567.
XX
PR   26-JAN-1990;   90US-0470888.
XX
PA   (BIOG-) BIOGEN INC.
XX
PI   Pasek MP,  Wirsler G,  Liu TR;
XX
DR   WPI; 1991-252613/34.
DR   N-PSDB; AAQ13242.
XX
PT   New C4 binding protein fusion proteins and DNA encoding them
PT   comprise assemblies of C4bp monomers linked to functional moiety,
PT   e.g. AZT, useful as delivery vehicles in diagnosis and therapy
XX
PS   Example 1; Fig 1; 105pp; English.
XX
CC   This sequence was deduced from human hepatocyte (Hep G2) cDNA
CC   obtained following PCR amplification. The protein is a monomer
CC   containing 8 SCRs. Each SCR forms a looped domain due to the
CC   presence of two intradomain disulphide bonds (only the disulphide
CC   bonds of SCR8 are labelled in the Features Table). Within each SCR,
CC   the first cysteine residue bonds with the third and the second
CC   cysteine residue bonds with the fourth. This secondary structure is
CC   responsible for the conformational flexibility of the C4bp monomer.
CC   The invention covers fusion proteins in which the monomer sequence,
CC   or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(S)
CC   is fused to the C terminal of a protein such as a viral receptor,
```

CC cell ligand, a bacterial, viral or parasitic immunogen, enzyme,
CC cytokine, toxin, etc. See also AAQ18243.51.
XX
SQ Sequence 581 AA;

Query Match 26.8%; Score 199; DB 12; Length 581;
Best Local Similarity 42.3%; Pred. No. 1.2e-11;
Matches 41; Conservative 20; Mismatches 60; Indels 6; Gaps 3;

Qy 2 ISCGSPPPIINGRISYYSTPIAVGTVIRYSC SGTPFRLIGEKSLLCITKDKVDG 61
Db 156 VKCKPPPDLRNGKHSGEDN-YAYGFSVTYSCDPRFSLIGHASISCTVENETIGVWRPSPP 216

Qy 62 PTLYENMYSSCPEPIVPGGYKIRGSTPYRDGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 217 PCEKI TCRKPDVSHGEMVSGFGPIYNYKDTIVFKCQKGFVLRGSSVIHCDADSKWN 272

Qy 121 PTRLPTC 127
Db 273 QSP PAC 278

RESULT 26
AAW73447
ID AAW73447 standard; protein; 778 AA.
XX
AC AAW73447;
XX
DT 29-JAN-1999 (first entry)
XX
DE Amino acid sequence of the soluble complement receptor 1 (sCR1).
XX
KW Human soluble complement receptor 1; sCR1; T cell; B cell;
KW mediated immune response inhibition; tissue rejection; gene therapy;
KW dystrophin; inflammatory response; interferon gamma secretory response;
KW autoimmune response; neurological response; Alzheimer's disease;
KW Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;
KW rheumatoid arthritis; myasthenia gravis; epidermis bullosa;
KW Hashimoto's disease.
XX
OS Homo sapiens.
XX
PN WO9845430-A1.
XX
PD 15-OCT-1998.
XX
PF 06-APR-1998; 98WO-GB01012.
XX
PR 05-APR-1997; 97GB-0006950.
XX
PA (ANNE/) ANNENKOV A.
PA (CHER/) CHERNAJOVSKY Y.
XX
PI Annenkov A, Chernajovsky Y
XX
DR WPI; 1998-508450/44.
XX
PT Fragment of soluble human complement receptor 1 useful for
PT treating T cell or B cell mediated immune responses e.g.
PT inflammatory responses such as rheumatoid arthritis
XX
PS Disclosure; Fig 1; 54pp; English.
XX
CC This is an amino acid sequence of the human soluble complement
CC receptor 1 (sCR1), useful in the treatment of T cell or B cell
CC mediated immune responses. It is used to inhibit a T cell or
CC B cell mediated immune response to prevent immune response mediated
CC tissue rejection and destruction or clearance or inactivation of an
CC expressed protein especially from cells that have been treated by gene
CC therapy to express the protein, e.g. dystrophin. The protein can also
CC be used to inhibit a T cell or B cell mediated inflammatory response,
CC an interferon gamma secretory response, autoimmune response or
CC neurological response, e.g. Alzheimer's or Parkinson's disease or
CC multiple sclerosis. Also the protein can be used to treat systemic
CC lupus erythematosus, rheumatoid arthritis, myasthenia gravis,
CC epidermis bullosa or Hashimoto's disease.
XX
SQ Sequence 778 AA;

Query Match 26.5%; Score 199; DB 19; Length 778;
Best Local Similarity 38.0%; Pred. No. 1.7e-11;
Matches 52; Conservative 14; Mismatches 59; Indels 12; Gaps 6;

Qy 2 ISCGSPPPIINGRISYYSTPIAVGTVIRYSC SG TPFRLIGEKSLLCITKDKVDG 94
Db 619 IPCGLPPTIANGDFISTNRENFHYGSVVIYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 678

Qy 95 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRDGDSVTFACKTNFSMNGNKSVWC 153
Db 679 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 735

Qy 154 QANNMWGPTRLPTCVSV 170
Db 736 QALNKWEPELPSCSRV 751

RESULT 27
AAR47156
ID AAR47156 standard; protein; 133 AA.
XX
AC AAR47156;
XX
DT 18-JUL-1994 (first entry)
XX
DE Sequence of soluble complement receptor type 1 (CR1) derivative
DE which comprises internal fragment MH122 S258 of CR1.
XX
KW Complement receptor type 1; CR1; short consensus repeat;
KW long homologous repeat; domain; inflammation; therapy.
XX
OS Synthetic.
XX
PN WO9400571-A.
XX
PD 06-JAN-1994.
XX
PF 16-JUN-1993; 93WO-GB01282.
XX
PR 24-JUN-1992; 92GB-0013376.
PR 01-MAR-1993; 93GB-0004057.
XX
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Dodd I, Freeman AM, Mossakowska DEI, Smith RAG;
XX
DR WPI; 1994-026208/03.
XX
PT New soluble complement receptor type 1 derivs. - used for
PT treating disease or disorder associated with inflammation or
PT inappropriate complement activation
XX
PS Claim 13; Page 54-55; 65pp; English.
XX
CC CR1 is composed of 30 short consensus repeats (SCRs) that each
CC contain around 60-70 AAs. It is further arranged as 4 long
CC homologous repeats (LHRs) of 7 SCRs each. Following a leader
CC sequence, the CR1 molecule consists of the N-terminal LHR A, the
CC next two repeats, LHR B and LHR-C, and the most C-terminal LHR D,
CC followed by 2 additional SCRs, a 25 residue putative transmembrane
CC region and a 44 residue cytoplasmic tail. Based on the mature CR1
CC molecule, having a predicted N-terminal Gln, designated residue 1,
CC the first four SCR domains of LHR A are 2-58, 63-120, 125-191 and
CC 197-252, of mature CR1. Soluble fragments of CR1 which corresp. to
CC part of CR1 possess functional complement inhibitory, including
CC anti haemolytic, activity.
XX

```
SQ   Sequence   133 AA;

  Query Match             26.4%;  Score 198;  DB 15;  Length 133;
  Best Local Similarity   38.1%;  Pred. No. 2.4e-12;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGR-ISYYSTPIAVGTVIRYSC--SG--TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        :|:| | |   ||    | |:|| |: | : |   |
Db       3 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 62

Qy      55 TWDKPAPKCEYFNKYSSCPEPTVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   | | | |  :  : :  :   : | | |:  | | | : | |
Db      63 IWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 119

Qy     114 QANNMWGPTRLPTC 127
           || | | |  ||:|
Db     120 QALNKWEP ELPSC 132


RESULT 28
AAR47154
ID   AAR47154 standard; protein; 254 AA.
XX
AC   AAR47154;
XX
DT   18-JUL-1994  (first entry)
XX
DE   Sequence of soluble complement receptor type 1 (CR1) derivative
DE   which comprises the N terminal fragment MQ1-S253 of CR1.
XX
KW   Complement receptor type 1; CR1; short consensus repeat;
KW   long homologous repeat; domain; inflammation; therapy.
XX
OS   Synthetic.
XX
PN   WO9400571 A.
XX
PD   06-JAN-1994.
XX
PF   16-JUN-1993;   93WO-GB01282.
XX
PR   24-JUN-1992;   92GB-0013476.
PR   01-MAR-1993;   93GB-0004057.
XX
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Dodd I,  Freeman AM,  Mossakowska DEI,  Smith RAG;
XX
DR   WPI; 1994-026208/03.
XX
PT   New soluble complement receptor type 1 derivs. - used for
PT   treating disease or disorder associated with inflammation or
PT   inappropriate complement activation
XX
PS   Claim 13; Page 51-52; 65pp; English.
XX
CC   CR1 is composed of 30 short consensus repeats (SCRs) that each
CC   contain around 60-70 AAs. It is further arranged as 4 long
CC   homologous repeats (LHRs) of 7 SCRs each. Following a leader
CC   sequence, the CR1 molecule consists of the N terminal LHR-A, the
CC   next two repeats, LHR-B and LHR-C, and the most C-terminal LHR-D,
CC   followed by 2 additional SCRs, a 25 residue putative transmembrane
CC   region and a 43 residue cytoplasmic tail. Based on the mature CR1
CC   molecule, having a predicted N-terminal Glu, designated residue 1,
CC   the first four SCR domains of LHR-A are 2-58, 63-120, 125-191 and
CC   197-252, of mature CR1. Soluble fragments of CR1 which corresp. to
CC   part of CR1 possess functional complement inhibitory, including
CC   anti haemolytic, activity.
XX
SQ   Sequence   254 AA;

  Query Match             26.4%;  Score 198;  DB 15;  Length 254;
  Best Local Similarity   38.1%;  Pred. No. 5.3e-12;
  Matches   51;  Conservative   14;  Mismatches   57;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGR-ISYYSTPIAVGTVIRYSC--SG--TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        :|:| | |   ||    | |:|| |: | : |   |
Db     124 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDQVG 183

Qy      55 TWDKPAPKCEYFNKYSSCPEPTVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   | | | |  :  : :  :   : | | |:  | | | : | |
Db     184 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTC 127
           || | | |  ||:|
Db     241 QALNKWEP ELPSC 253


RESULT 29
AAY55754
ID   AAY55754 standard; Protein; 450 AA
XX
AC   AAY55754;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR-C SCR fragment.
XX
KW   C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1985;   85US-0332865.
PR   06-DEC-1974;   74US-0350238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176532.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ij SH,
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-533357/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having antiinflammatory and
PT   cardiant activity -
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C3b/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR) C sequence.
```

```
XX
SQ   Sequence   450 AA;

  Query Match             26.4%;  Score 198;  DB 20;  Length 450;
  Best Local Similarity   35.5%;  Pred. No. 1.1e-11;
  Matches   50;  Conservative   15;  Mismatches   60;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPTILNGR-ISYYSTPIAVGTVIRYSCS      GTPRLIGEESLLCTIKDKVDG 54
           | || || | ||  ||        |:|: | |:        | |:|| | : | :  |
Db     124 IPCGLPPTIANGDFISTNRENFHYGSVVTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVG 183

Qy      55 TWDKPAPKCEYFNKYSSCPPPIVPGGYKIRGS-TPYRHGDSVIFACKTNFSMNGNKSVWC 113
            |  :|||:   ||      | | |  :  : :   : | | |:  | : | :  | | |
Db     184 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  |:|  |
Db     241 QALNKWEP-ELPSCSRV 256


RESULT 30
AAY55757
ID   AAY55757 standard; Protein; 453 AA.
XX
AC   AAY55757;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR-1 fragment.
XX
KW   C3b/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR.
XX
OS   Homo sapiens.
XX
PN   US5981481-A
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1984;   84US-0596865.
PR   30-DEC-1974;   74US-0536238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176562.
XX
PA   (JOHN-) UNIV JOHNS HOPKINS.
PA   (BRIG-) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Cosimo MI;  Wong WW;  Makrides SC;  Klickstein LB;  Fearon DT;  Ip SH;
PI   Marsh HC;  Carson CR;
XX
DR   WPI; 1999-633857/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having anti-inflammatory and
PT   cardiant activity
XX
PS   Disclosure; Fig 5B; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the human CR1 protein long homologous repeat (LHR) C fragment.
XX
SQ   Sequence   453 AA;

  Query Match             26.4%;  Score 198;  DB 20;  Length 453;
  Best Local Similarity   35.5%;  Pred. No. 1.1e-11;
  Matches   50;  Conservative   15;  Mismatches   60;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPTILNGR-ISYYSTPIAVGTVIRYSCS      GTPRLIGEESLLCTIKDKVDG 54
           | || || | ||  ||        |:|: | |:        | |:|| | : | :  |
Db     124 IPCGLPPTIANGDFISTNRENFHYGSVVTYRCNLGSRGRKVFELVGEPSIYCTSNDDQVG 183

Qy      55 TWDKPAPKCEYFNKYSSCPPPIVPGGYKIRGS-TPYRHGDSVIFACKTNFSMNGNKSVWC 113
            |  :|||:   ||      | | |  :  : :   : | | |:  | : | :  | | |
Db     184 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 240

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  |:|  |
Db     241 QALNKWEP-ELPSCSRV 256


RESULT 31
AAR29092
ID   AAR29092 standard; peptide; 481 AA.
XX
AC   AAR29092;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (amino acids 61-122 deleted) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR 1
FT   Region          389..448
FT                   /label= SCR 8
FT   Region          449..481
FT                   /label= SCR 9
FT                   /note= "TRUNCATED"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0304826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW-) UNIV WASHINGTON.
XX
PI   Atkinson JP;  Bonracade D;  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating autoimmune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed including this
```

```
CC   one in which the entire SCR 2 has been deleted. The variants have
CC   altered C3b- and C4b binding properties. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   481 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 481;
  Best Local Similarity   38.0%;  Pred. No. 1.9e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC---SG---TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  || |  |:|: | |   ||    | |:|| |: | : |   |
Db      61 IPCGLPPTIFNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 120

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWC 113
            |  |||:|   ||   |  | |  |  :  : : :   : | | |:  | | | : | |
Db     121 IWSGPAPQCIIPNK--CTIPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 177

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     178 QALNKWEP-ELPSCSRV 193


RESULT 32
AAR29091
ID   AAR29091 standard; peptide; 483 AA.
XX
AC   AAR29091;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (amino acids 1-60 deleted) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..62
FT                   /label= SCR-2
FT   Region          391..450
FT                   /label= SCR-8
FT   Region          451..483
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 24pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J Exp Med (1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
```

```
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed including this
CC   one in which the entire SCR-1 has been deleted. The variants have
CC   altered C3b- and C4b-binding properties. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   483 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 483;
  Best Local Similarity   38.0%;  Pred. No. 1.9e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC---SG---TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  || |  |:|: | |   ||    | |:|| |: | : |   |
Db      63 IPCGLPPTIFNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 122

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWC 113
            |  |||:|   ||   |  | |  |  :  : : :   : | | |:  | | | : | |
Db     123 IWSGPAPQCIIPNK--CTIPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 179

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     180 QALNKWEP-ELPSCSRV 195


RESULT 33
AAR28543
ID   AAR28543 standard; peptide; 543 AA
XX
AC   AAR28543;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (35E, 37Y) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 35
FT                   /note= "Gly substituted by Glu (SCR 8)"
FT   Misc-difference 37
FT                   /note= "Ser substituted by Tyr (SCR 8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
```

```
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analoques of "requlator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1 4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed. Positions 35 and
CC   37 of SCR-1 and the corresponding positions in SCR 8 have been
CC   identified as important in C4b binding. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match              26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity    38.0%;  Prod. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSITVAVGIVIRYSC--SG--IERLIGEKSLLCIIKDKVDG 54
           | || || | ||  ||         |:|: | |   ||    | |:|| |: | : |   |
Db     123 IPCGIPPLIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-IDYRHGDSVTFACKINFSMNGNKSVWC 113
           | .|||:|   ||      | | |  :  : :  :  : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGIIVSDNRSLPSLNEVVEFRCQPGFVMMGSPRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     240 QALNKWEP-ELPSCSRV 255


RESULT 34
AAR28544
ID   AAR28544 standard; peptide; 543 AA.
XX
AC   AAR28544;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (35Y) analoque.
XX
KW   short consensus repeat; requlator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc difference 35
FT                   /note= "Gly substituted by Glu (SCR-8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303825.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON
XX
PI   Atkinson JP, Hourcade D, Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity requlator protein analoques useful for
PT   treating auto;immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analoques of "requlator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed. Positions 35 and
CC   37 of SCR-1 and the corresponding positions in SCR 8 have been
CC   identified as important in C4b binding. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match              26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity    38.0%;  Prod. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSITVAVGIVIRYSC--SG--IERLIGEKSLLCIIKDKVDG 54
           | || || | ||  ||         |:|: | |   ||    | |:|| |: | : |   |
Db     123 IPCGIPPLIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-IDYRHGDSVTFACKINFSMNGNKSVWC 113
           | .|||:|   ||      | | |  :  : :  :  : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGIIVSDNRSLPSLNEVVEFRCQPGFVMMGSPRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     240 QALNKWEP-ELPSCSRV 255


RESULT 35
AAR28545
ID   AAR28545 standard; peptide; 543 AA.
XX
AC   AAR28545;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (37Y) analoque.
XX
KW   short consensus repeat; requlator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc difference 37
FT                   /note= "Ser substituted by Tyr (SCR-8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
```

```
PF   28 APR 1992;   92EP 0303826.
XX
PR   03 MAY-1991;   91US-0695514
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto:immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255 1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1 4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed. Positions 35 and
CC   37 of SCR 1 and the corresponding positions in SCR-8 have been
CC   identified as important in C4b binding. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12,  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC--SG--TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db     123 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | |||:|   ||   | | | |  :  : : :   : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     240 QALNKWEP-ELPSCSRV 255


RESULT 36
AAR28546
ID   AAR28546 standard; peptide; 543 AA.
XX
AC   AAR28546;
XX
DT   19 MAR 1993  (first entry)
XX
DE   CR1 4 (44T, 47D, 49L) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR 1
FT   Region          61..122
FT                   /label= SCR 2
FT   Region          451..510
FT                   /label= SCR 8
FT   Region          511..543
FT                   /label= SCR 9
FT                   /note= "TRUNCATED"
FT   Misc difference 44
FT                   /note= "Ile substituted by Thr (SCR-8)"
FT   Misc-difference 47
FT                   /note= "Lys substituted by Asp (SCR 8)"
FT   Misc-difference 49
FT                   /note= "Ser substituted by Leu (SCR 8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP   Hourcade D,  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto:immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255 1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-1 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-8. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC--SG--TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|:      ||    | |:|| |: | : |   |
Db     123 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGPKVFELVGEPSIYCTSNDDQVG 182

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | |||:|   ||   | | | |  :  :   :   : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || |   |  ||:|  |
Db     240 QALNKWEP-ELPSCSRV 255


RESULT 37
AAR28547
ID   AAR28547 standard; peptide; 543 AA
XX
AC   AAR28547;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (52S, 53S, 54P) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens
XX
```

```
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 57
FT                   /note= "Arg substituted by Val (SCR-8)"
FT   Misc-difference 59
FT                   /note= "Arg substituted by Lys (SCR-8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;  92EP-0303826.
XX
PR   03-MAY-1991;  91US-0695514.
XX
PA   (UNIW-) UNIV WASHINGTON.
XX
PI   Atkinson JP, Bourcade D, Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-1 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-8. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC   SG   TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db     123 IPCGLPPTIINGDFISINRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRHGDSVIFACKTNFSMNGNKSVWC 113
             | |:| |  ||  | | : : : :  : | | |: | | | : | |
Db     183 IWSGPAPQCIIPNK  CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           |: | | |  |:|  |
Db     240 QALNKWEP ELPSCSRV 255


RESULT 58
AAR28548
ID   AAR28548 standard; peptide; 543 AA.
XX
AC   AAR28548;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (57V, 59K) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 57
FT                   /note= "Arg substituted by Val (SCR-8)"
FT   Misc-difference 59
FT                   /note= "Arg substituted by Lys (SCR-8)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;  92EP-0303826.
XX
PR   03-MAY-1991;  91US-0695514.
XX
PA   (UNIW-) UNIV WASHINGTON.
XX
PI   Atkinson JP, Bourcade D, Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-1 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-8. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPILNGR-ISYYSTPIAVGTVIRYSC   SG   TFRLIGEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |   ||    | |:|| |: | : |   |
Db     123 IPCGLPPTIINGDFISINRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 IWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS TPYRHGDSVIFACKTNFSMNGNKSVWC 113
             | |:| |  ||  | | : : : :  : | | |: | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239
```

```
Qy      114 QANNMWGPTRLPTCVSV 130
            || | | |  ||:|  |
Db      240 QALNKWEP-ELPSCSRV 255


RESULT 39
AAR28549
ID   AAR28549 standard; peptide; 543 AA.
XX
AC   AAR28549;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1 4 (64K, 65T) analoque.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C4b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc difference 64
FT                   /note= "Arg substituted by Lys (SCR-9)"
FT   Misc difference 65
FT                   /note= "Asn substituted by Thr (SCR 9)"
XX
PN   EP512733 A.
XX
PD   11 NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03 MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto:immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 24pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J.Exp.Med.(1988)
CC   168:1255 1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analoques of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C4b  and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy        2 ISCGSPPLINGR-ISYYSTPIAVGTVIRYSC  -SG   TFRLIGEKSLLCITKDKVDG 54
            | ||  | | ||  ||        |:|:  |   ||    | |:|| |: | : |   |
Db      223 IPCGLPPIITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 282

Qy       55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIEGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   | | |  |  : : : :  : | | |:  | | : | | |
Db      283 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 339

Qy      114 QANNMWGPTRLPTCVSV 130
            || | |    ||:|  |
Db      240 QALNKWEP-ELPSCSRV 255


RESULT 40
AAR28550
ID   AAR28550 standard; peptide; 543 AA.
XX
AC   AAR28550;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (64K) analoque.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 64
FT                   /note= "Arg substituted by Lys (SCR-9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992
XX
PF   28-APR-1992.   92EP-0303826.
XX
PR   03-MAY 1991:   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M.
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator prote.n ana.oques - useful for
PT   treating auto:immune diseases, to suppress transplant rejection
PT   for diagnosis etc.
XX
PS   Claim 11, Fig 2 and R11810; 24pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255 1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analoques of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full-length CR1 amino acid sequence having GENESEQ
```

```
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPEIINGR ISYYSIPIAVGIVIFYSC   SG   TFFLICEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |    ||    | |:|| | : | : |   |
Db     123 IPCGLPPIIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 TWIKPAPKCEYENKYSSCPEPIVPGGYKIRGS TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | ||| |   ||   | | | |  :  : : :   : | | |: | | | : | |
Db     183 IWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           ||   | |  |||:
Db     240 QALNKWEP ELPSCSRV 255


RESULT 41
AAR28551
ID   AAR28551 standard; peptide; 543 AA.
XX
AC   AAR28551;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (6G1) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 64
FT                   /note= "Arg substituted by Lys (SCR 9)"
FT   Misc-difference 65
FT                   /note= "Asn substituted by Thr (SCR 9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP;  Hourcade D;  Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C3b and C4b binding are substituted by amino acids from
CC   the corresponding positions in SCR 9. The specification does not
CC   contain the CR1 4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPEIINGR ISYYSIPIAVGIVIFYSC   SG   TFFLICEKSLLCITKDKVDG 54
           | || || | ||  ||        |:|: | |    ||    | |:|| | : | : |   |
Db     123 IPCGLPPIIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy      55 TWIKPAPKCEYENKYSSCPEPIVPGGYKIRGS TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | ||| |   ||   | | | |  :  : : :   : | | |: | | | : | |
Db     183 IWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           ||   | |  |||:
Db     240 QALNKWEP ELPSCSRV 255


RESULT 42
AAR28552
ID   AAR28552 standard; peptide; 543 AA.
XX
AC   AAR28552;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (78T, 79D) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 78
FT                   /note= "Lys substituted by Thr (SCR 9)"
FT   Misc-difference 79
FT                   /note= "Gly substituted by Asp (SCR 9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP;  Hourcade D;  Krych M;
XX
DR   WPI; 1992-375009/46.
```

```
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR 2 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGR-ISYYSTPIAVGTVIRYSC---SG---TPRLIGEKSLLCITKDKVDC 54
           | || || | ||  ||         |:|: | |    ||    | |:|| |: | : |    |
Db     123 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVC 182

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | | |  |  : : : :   : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK-  CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     240 QALNKWEP ELPSCSRV 255


RESULT 43
AAR28553
ID   AAR28553 standard; peptide; 543 AA.
XX
AC   AAR28553;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (85R, 87N) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc difference 85
FT                   /note= "Gln substituted by Arg (SCR-9)"
FT   Misc difference 87
FT                   /note= "Lys substituted by Asn (SCR-9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;   92EP-0303826.
XX
PR   03-MAY-1991;   91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP,  Hourcade D,  Krych M;
XX
DR   W-1; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and R11810; 23pp; English.
XX
CC   The cDNA clone designated CR1-4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGR-ISYYSTPIAVGTVIRYSC---SG---TPRLIGEKSLLCITKDKVDC 54
           | || || | ||  ||         |:|: | |    ||    | |:|| |: | : |    |
Db     123 IPCGLPPTITNGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVC 182

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYK RGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             |  |||:|   ||   |  | | |  |  : : : :   : | | |:  | | | : | |
Db     183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPVFVMKGPRRVKC 239

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     240 QALNKWEP-ELPSCSRV 255


RESULT 44
AAR28554
ID   AAR28554 standard; peptide; 543 AA.
XX
AC   AAR28554;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (92', 94H) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
```

```
FT   Misc-difference 92
FT                   /note= "Lys substituted by Thr (SCR 9)"
FT   Misc-difference 94
FT                   /note= "Tyr substituted by His (SCR 9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;    92EP-0304826.
XX
PR   03-MAY-1991;    91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP, Bourcade D, Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and 811810; 23pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy        2 ISCGSPPPIINGR ISYYSTPIAVGTVIRYSC   SG-- TFRLIGEKSLLCITKDKVDG 54
            | || || | || ||        |:|: | |   ||    | |:|| |: | : |   |
Db      123 IPCGLPPTIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy       55 IWDKPAPKCIYYNKYSSCPEPIVPGGYKIRGS TPYRHGDSVIFACKINFSMNGNKSVWC 113
             | :|:|   ||      | | |  : : :   : | | |:  | | | : | |
Db      183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 239

Qy      114 QANNMWGPTRLPTCVSV 130
            || | | |  |:|: |
Db      240 QAENKWEP-ELPSCSRV 255


RESULT 4
AAR28555
ID   AAR28555 standard; peptide; 543 AA.
XX
AC   AAR28555;
XX
DT   19-MAR-1993  (first entry)
XX
DE   CR1-4 (92I) analogue.
XX
KW   short consensus repeat; regulator of complement activation;
KW   C3b binding; C4b binding; human complement type 1 receptor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..60
FT                   /label= SCR-1
FT   Region          61..122
FT                   /label= SCR-2
FT   Region          451..510
FT                   /label= SCR-8
FT   Region          511..543
FT                   /label= SCR-9
FT                   /note= "TRUNCATED"
FT   Misc-difference 92
FT                   /note= "Lys substituted by Thr (SCR 9)"
XX
PN   EP512733-A.
XX
PD   11-NOV-1992.
XX
PF   28-APR-1992;    92EP-0304826.
XX
PR   03-MAY-1991;    91US-0695514.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Atkinson JP, Bourcade D, Krych M;
XX
DR   WPI; 1992-375009/46.
XX
PT   Complement activity regulator protein analogues - useful for
PT   treating auto-immune diseases, to suppress transplant rejection,
PT   for diagnosis etc.
XX
PS   Claim 11; Fig 2 and 811810; 23pp; English.
XX
CC   The cDNA clone designated CR1 4 was described in J.Exp.Med.(1988)
CC   168:1255-1270. It encodes the first 8 and a half amino terminal
CC   SCRs of CR1. The invention concerns analogues of "regulator of
CC   complement activation" proteins or truncated, hybrid or recombinant
CC   forms of them. CR1-4 is a preferred truncated form and a number of
CC   specified substitution variants of it are claimed in which certain
CC   positions in SCR-2 which have been identified as important for the
CC   degree of C3b- and C4b-binding are substituted by amino acids from
CC   the corresponding positions in SCR-9. The specification does not
CC   contain the CR1-4 sequence; the sequence given here was constructed
CC   from the full length CR1 amino acid sequence having GENESEQ
CC   accession number AAR11810 and descriptions in the disclosure.
XX
SQ   Sequence   543 AA;

  Query Match             26.1%;  Score 196;  DB 13;  Length 543;
  Best Local Similarity   38.0%;  Pred. No. 2.2e-11;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy        2 ISCGSPPPIINGR ISYYSTPIAVGTVIRYSC   SG-- TFRLIGEKSLLCITKDKVDG 54
            | || || | || ||        |:|: | |   ||    | |:|| |: | : |   |
Db      123 IPCGLPPTIINGDFISTNRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 182

Qy       55 IWDKPAPKCIYYNKYSSCPEPIVPGGYKIRGS TPYRHGDSVIFACKINFSMNGNKSVWC 113
             | :|:|   ||      | | |  : : :   : | | |:  | | | : | |
Db      183 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 239

Qy      114 QANNMWGPTRLPTCVSV 130
            || | | |  |:|: |
Db      240 QAENKWEP-ELPSCSRV 255


Search completed: November  6, 2002, 16:06:28
Job time : 32.8996 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 14.4461 Seconds
(without alignments)
891.311 Million cell updates/sec

Title: US-09-834-309-4
Perfect score: 751
Sequence: 1 GISCGSPPPILNGRISYYST ... ANNMWGPTPLPTCVSVFPLE 134

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	751	100.0	1091	1	PLD009	complement C3d/Eps
2	466.5	62.1	364	2	B45900	complement C3d/Eps
3	466.5	62.1	676	2	A45900	complement C3b rec
4	466.5	62.1	1025	1	A43526	complement C3d/Eps
5	220	29.3	2489	2	I73012	complement C3b/C4b
6	218	29.0	2014	2	I36046	complement recepto
7	215	28.6	264	1	WMVXSP	apolipoprotein H h
8	210	28.0	264	1	C36838	complement control
9	210	28.0	264	2	I28450	hypothetical prote
10	208	27.7	264	2	B72152	B18L protein - var
11	203.5	27.1	579	2	A56740	sperm egg recognit
12	202.5	27.0	597	1	S53711	C4BP alpha chain p
13	201	26.8	482	2	A34924	complement C3b/C4b
14	199	26.5	597	1	NBHUC4	C4b-binding protei
15	197.5	26.3	610	1	I46001	C4b-binding protei
16	184	24.5	497	2	JC2054	complement regulat
17	182	24.2	568	2	S57953	C4BP protein alpha
18	180.5	24.0	440	2	A43519	complement recepto
19	180.5	24.0	1231	1	NBHUH	complement factor
20	178.5	23.8	469	1	NBMSC4	C4b-binding protei
21	178.5	23.8	1234	1	NBMSH	complement factor
22	165.5	22.0	808	2	D35069	complement factor
23	165	22.0	661	1	KFHUI3	coagulation factor
24	164	21.8	2044	2	T18524	scavenger receptor
25	162	21.6	343	2	G35070	apolipoprotein H-r
26	160	21.3	362	2	JC5194	membrane cofactor
27	160	21.3	369	2	JC5138	membrane cofactor
28	159	21.2	377	2	I54479	membrane cofactor
29	159	21.2	384	2	S01896	membrane cofactor
30	158.5	21.1	340	2	I56234	decay-accelerating
31	158.5	21.1	381	1	B26359	decay-accelerating
32	158.5	21.1	440	2	A26359	decay-accelerating
33	155	20.6	349	2	G02913	sperm CD46 - human
34	155	20.6	369	2	I57998	membrane cofactor
35	150	20.0	302	1	WMBE1E	secretory compleme
36	150	20.0	360	1	WMBE2E	membrane bound com
37	149	19.8	668	2	A46013	coagulation factor
38	148	19.7	768	2	A42755	P-selectin precurs
39	147	19.6	330	2	I56100	complement factor
40	146.5	19.5	768	2	I53821	P-selectin - rat
41	142	18.9	1019	2	A38738	coagulation factor
42	141	18.8	252	2	A34877	C4b-binding protei
43	140	18.6	152	2	A53274	complement factor
44	140	18.6	258	2	S57960	C4BP protein beta
45	140	18.6	752	1	C2HU	complement C2 prec
46	137	18.2	345	1	NBMS	apolipoprotein H p
47	135	18.0	360	2	T42921	complement control
48	134.5	17.9	270	2	I37278	complement factor
49	134	17.8	297	1	NBRT	apolipoprotein H p
50	134	17.8	763	2	I50807	complement factor
51	133	17.7	764	1	BBHU	complement factor
52	132	17.6	345	1	JN0465	apolipoprotein H p
53	131	17.4	202	1	A44247	C4b binding protei
54	131	17.4	669	2	S65551	factor H - bovine
55	129.5	17.2	702	2	T16842	hypothetical prote
56	128	17.0	560	2	T16843	hypothetical prote
57	127	16.9	345	1	NBHU	apolipoprotein H p
58	127	16.9	345	1	NBBO	apolipoprotein H p
59	127	16.9	646	2	JN0473	P-selectin precurs
60	125	16.6	868	2	T20239	hypothetical prote
61	125	16.6	1827	2	T34388	hypothetical prote
62	124.5	16.6	977	2	I52657	seizure-related pr
63	123.5	16.4	612	2	B42755	E-selectin precurs
64	123	16.4	934	1	A34372	complement C6 prec
65	122	16.2	761	1	BBMS	complement factor
66	120.5	16.0	551	2	I46709	endothelial leuko
67	120.5	16.0	760	1	C2MS	classical compleme
68	120	16.0	330	2	I55975	X/Y protein - mous
69	120	16.0	610	2	A35046	E-selectin precurs
70	118	15.7	331	2	A45222	complement factor
71	118	15.7	482	2	JC5092	E-selectin - pig
72	116.5	15.5	1797	2	T21889	hypothetical prote
73	116.5	15.5	1805	2	T21888	hypothetical prote
74	116	15.4	303	2	H35068	apolipoprotein H r
75	114	15.2	452	2	A35068	complement factor
76	114	15.2	830	2	A30359	P-selectin precurs
77	113	15.0	570	2	T46261	hypothetical prote
78	111.5	14.8	449	1	NBHUHS	complement factor
79	111.5	14.8	1506	2	I30886	integumentary muci
80	111	14.8	747	2	I51579	complement factor
81	110.5	14.7	843	1	A27340	complement C7 prec
82	110	14.6	485	2	G36772	E-selectin - bovin
83	106.5	14.2	1053	2	S46199	probable complemen
84	102.5	13.6	317	2	A38493	pS/Zr protein - va
85	101.5	13.5	686	1	A59271	Ra reactive factor
86	98.5	13.1	317	2	T37442	HIV gp42,pS/Zr pro
87	97.5	13.0	317	2	T28605	hypothetical prote
88	94.5	12.6	317	2	JQ1799	B5R protein precur
89	93	12.4	347	1	HPRT	haptoglobin precur
90	91.5	12.2	317	2	F72172	B7R protein - vari
91	91.5	12.2	317	2	G36855	B7R protein - vari
92	90.5	12.1	317	2	D42526	B5R protein - vacc
93	86	11.5	699	1	I54763	Ra-reactive factor
94	85.5	11.4	926	1	OPPGIT	iodide peroxidase
95	85	11.3	198	2	I46002	C4BP beta chain
96	85	11.3	347	1	HPMS	haptoglobin precur
97	85	11.3	633	2	T24898	hypothetical prote
98	84	11.2	370	2	S22124	E-selectin precurs
99	83.5	11.1	406	1	HPHU2	haptoglobin precur
100	83	11.1	933	1	OPHUIT	iodide peroxidase

ALIGNMENTS

RESULT 1
JL0028
complement C3d/Epstein-Barr virus receptor precursor - human
N;Alternate names: complement receptor 2; CR2/CD21
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1992 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1993
C;Accession: JL0028; A45958; A32036; A24419; B24419; S24419; S24434; F24419; F24439; P12
R;Weis, J.J.; Tedder, T.F.; Fearon, D.T.
J. Exp. Med. 167, 1047-1066, 1988
A;Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus
A;Reference number: JL0028; MUID:88171282
A;Accession: JL0028
A;Molecule type: mRNA
A;Residues: 1-1091 <WEI>
A;Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542 hav
R;Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987
A;Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (com
A;Reference number: A45958; MUID:88097454
A;Accession: A45958
A;Molecule type: mRNA
A;Residues: 1-476,'G',477-614,'R',616-669,'R',671-816,'NCSAFVILKAWILFRAF',835-840,'L',84
A;Cross-references: GB:J03565; NID:g181919; PIDN:AAA35784.1; PID:g181920
R;Fujisaku, A.; Harley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Bolers, V.M.
J. Biol. Chem. 264, 2118-2125, 1989
A;Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus rece
A;Reference number: A32036; MUID:89123277
A;Accession: A32036
A;Molecule type: mRNA
A;Residues: 1-476,'G',477-658,718-1050,'T',1052-1060,'E',1062-1091 <FUJ>
A;Cross-references: GB:J04463
R;Weis, J.J.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Kops,
Proc. Natl. Acad. Sci. U.S.A. 83, 5639-5643, 1986
A;Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor
of complement.
A;Reference number: A94114; MUID:86287411
A;Accession: A24419
A;Molecule type: protein
A;Residues: 226-230,'XIIQ';257-267;332-341;583-591,'Q',594,'D',595-596;728-735 <WE2>
A;Experimental sources: B lymphoblastoid cell lines SB and Raji
C;Genetics:
A;Gene: GDB:CR2
A;Cross-references: GDB:119802; OMIM:120650
A;Map position: 1q32-1q32
C;Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat ho
C;Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-1091/Product: complement receptor 2 (16 repeat form) #status predicted <MAT>
F;21-558/718-1091/Product: complement receptor 2 (15 repeat form) #status predicted <MAT
F;23-86/Domain: complement factor H repeat homology <FH01>
F;91-146/Domain: complement factor H repeat homology <FH02>
F;154-210/Domain: complement factor H repeat homology <FH03>
F;215-271/Domain: complement factor H repeat homology <FH04>
F;276-342/Domain: complement factor H repeat homology <FH05>
F;351-406/Domain: complement factor H repeat homology <FH06>
F;410-465/Domain: complement factor H repeat homology <FH07>
F;470-521/Domain: complement factor H repeat homology <FH08>
F;526-582/Domain: complement factor H repeat homology <FH09>
F;601-656/Domain: complement factor H repeat homology <FH10>
F;660-716/Domain: complement factor H repeat homology <FH11>
F;720-777/Domain: complement factor H repeat homology <FH12>
F;777-837/Domain: complement factor H repeat homology <FH13>
F;846-901/Domain: complement factor H repeat homology <FH14>
F;909-965/Domain: complement factor H repeat homology <FH15>
F;970-1026/Domain: complement factor H repeat homology <FH16>
F;1034-1056/Domain: transmembrane #status predicted <TMM>
F;1057-1091/Domain: intracellular #status predicted <INT>
F;121,127,294,472,622,698,858-881,919/Binding site: carbohydrate (Asn) (covalent) #statu

Query Match 100.0%; Score 751; DB 1; Length 1091;
Best Local Similarity 100.0%; Pred. No. 5.1e-64;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GISCGSPPPTLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 60
Db 23 GISCGSPPPTLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 82

Qy 61 PKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 83 PKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 142

Qy 121 PTRLPTCVSVPPLE 134
Db 143 PTRLPTCVSVPPLE 156

RESULT 2
B45900
complement C3d/Epstein-Barr virus receptor precursor alternative splice form - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45900; C45900
R;Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A;Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A;Reference number: A45900; MUID:90229754
A;Accession: B45900
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-464 <KUR>
A;Accession: C45900
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-19,'E',84-363 <KU2>
C;Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat
F;23-81/Domain: complement factor H repeat homology <FH1>
F;86-145/Domain: complement factor H repeat homology <FH2>
F;154-210/Domain: complement factor H repeat homology <FH3>
F;218-274/Domain: complement factor H repeat homology <FH4>
F;279-335/Domain: complement factor H repeat homology <FH5>

Query Match 62.1%; Score 466.5; DB 2; Length 363;
Best Local Similarity 61.2%; Pred. No. 4.1e-47;
Matches 82; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

Qy 2 ISCGSPPPTLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
Db 84 ISCGYPPEVKNAFKFYYSLPIVPGTVLEYTCEPGYRLIGEKAIFCISENQVHAIWDKAPP 143

Qy 62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 144 PCESVNKTSCSSPVIPGSEMNAGSKAFFKDDSVIEYKCAPGMRGSKIVWCQANEMWG 203

Qy 121 PTRLPTCVSVPPLE 134
Db 204 PTALPVCESDFPLE 217

RESULT 3
A45900
complement C3b receptor type 2 long form precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C;Accession: A45900; I48306
R;Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A;Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A;Reference number: A45900; MUID:90229754
A;Accession: A45900
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-476 <KUR>
A;Cross-references: GB:M86470

A;Experimental source: clone 41.1
R;Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.
J. Exp. Med. 181, 151-159, 1995
A;Title: Mouse complement regulatory protein Crry/p65 uses the specific mechanisms of bo
A;Reference number: I48306; MUID:95105691
A;Accession: I48406
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 21-467 <RES>
A;Cross-references: EMBL:U17128; NID:g595980; PIDN:AAA78271.1; PID:g595982
C;Genetics:
A;Gene: Cr2
A;Introns: 80/1; 113/2; 342/1; 2754; 3337
C;Superfamily: complement factor H repeat homology
F;22-78/Domain: complement factor H repeat homology <FH01>
F;83-140/Domain: complement factor H repeat homology <FH02>
F;145-211/Domain: complement factor H repeat homology <FH03>
F;217-272/Domain: complement factor H repeat homology <FH04>
F;276-331/Domain: complement factor H repeat homology <FH05>
F;336-394/Domain: complement factor H repeat homology <FH06>
F;399-458/Domain: complement factor H repeat homology <FH07>
F;467-523/Domain: complement factor H repeat homology <FH08>
F;531-587/Domain: complement factor H repeat homology <FH09>
F;592-648/Domain: complement factor H repeat homology <FH10>

Query Match 62.1%; Score 466.5; DB 2; Length 676;
Best Local Similarity 61.2%; Pred. No. 6.2e-37;
Matches 82; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

Qy 2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLICITKDKVDGTWDKPAP 61
Db 497 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 456

Qy 62 KCEYPNKYSSCPEPIVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 457 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 516

Qy 121 PTRLPTCVSVFPLE 134
Db 517 PTALPVCESDFPLE 530

RESULT 4
A43526
complement C3d/Epstein-Barr virus receptor 2 precursor - mouse
N;Alternate names: complement receptor type 2
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A43526; A43538; A32215; A45802; B32215
R;Finger oth, J.D.
J. Immunol. 144, 3458-3467, 1990
A;Title: Comparative structure and evolution of murine CR2. The homolog of the human C3d
A;Reference number: A43526; MUID:90229735
A;Accession: A43526
A;Molecule type: mRNA
A;Residues: 1-1025 <FIN>
A;Cross-references: GB:M35684; EMBL:J04153; NID:g192687; PIDN:AAA37448.1; PID:g192688
R;Molina, H.; Kinoshita, T.; Inoue, K.; Carel, J.C.; Holers, V.M.
J. Immunol. 145, 2974-2983, 1990
A;Title: A molecular and immunochemical characterization of mouse CR2. Evidence for a si
A;Reference number: A43538; MUID:91010789
A;Accession: A43538
A;Molecule type: mRNA
A;Residues: 12-305,'T',307-519,'A',521-1025 <MOL>
A;Cross-references: GB:M61132; NID:g192692; PIDN:AAA63295.1; PID:g192693
R;Finger oth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989
A;Title: Identification of murine complement receptor type 2.
A;Reference number: A32215; MUID:89098890
A;Accession: A32215
A;Molecule type: mRNA
A;Residues: 343-431;991-1025 <FI2>
A;Cross-references: GB:J04153

R;Kurtz, C.B.; Paul, M.S.; Aegerter, M.; Weis, J.J.; Weis, J.H.
J. Immunol. 143, 2058-2067, 1989
A;Title: Murine complement receptor gene family. Identification and characterization
A;Reference number: A45802; MUID:89381350
A;Accession: A45802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'E',100-101,292-961,964-1025 <KUR>
A;Cross-references: GB:M29281; NID:g192685; PIDN:AAA37447.1; PID:g387141
A;Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for resid
A;Note: the authors translated the codon CAC for residue 727 as Asn.
C;Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat
C;Keywords: alternative splicing; duplication; glycoprotein; receptor; transmembrane
F;1-11/Domain: signal sequence #status predicted <SIG>
F;12-1025/Product: complement C3d/Epstein-Barr virus receptor 2 (15-repeat form) #sta
F;12-973/Domain: extracellular #status predicted <EXT>
F;14-73/Domain: complement factor H repeat homology <FH01>
F;82-138/Domain: complement factor H repeat homology <FH02>
F;146-202/Domain: complement factor H repeat homology <FH03>
F;207-263/Domain: complement factor H repeat homology <FH04>
F;268-334/Domain: complement factor H repeat homology <FH05>
F;343-398/Domain: complement factor H repeat homology <FH06>
F;402-458/Domain: complement factor H repeat homology <FH07>
F;463-514/Domain: complement factor H repeat homology <FH08>
F;519-585/Domain: complement factor H repeat homology <FH09>
F;594-649/Domain: complement factor H repeat homology <FH10>
F;654-704/Domain: complement factor H repeat homology <FH12>
F;709-769/Domain: complement factor H repeat homology <FH13>
F;778-833/Domain: complement factor H repeat homology <FH14>
F;841-897/Domain: complement factor H repeat homology <FH15>
F;902-958/Domain: complement factor H repeat homology <FH16>
F;968-989/Domain: transmembrane #status predicted <TMM>
F;990-1025/Domain: intracellular #status predicted <INT>

Query Match 62.1%; Score 466.5; DB 1; Length 1025;
Best Local Similarity 61.2%; Pred. No. 9.7e-37;
Matches 82; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

Qy 2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLICITKDKVDGTWDKPAP 61
Db 12 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 71

Qy 62 KCEYPNKYSSCPEPIVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 72 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 131

Qy 121 PTRLPTCVSVFPLE 134
Db 132 PTALPVCESDFPLE 145

RESULT 5
I73012
complement C3b/C4b receptor, membrane-bound form precursor - human
N;Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); sur
N;Contains: complement C3b/C4b receptor, secreted form
C;Species: Homo sapiens (man)
C;Date: 02-Jul 1996 #sequence_revision 24-Nov-1999 #text_change 21-Jul 2000
C;Accession: I73012; I56203; A47602; S33291; S03843; A28507; A24748; B24748; C24748
R;Vik, D.P.; Wong, W.W.
J. Immunol. 15?, 6214-6224, 1993
A;Title: Structure of the gene for the F allele of complement receptor type 1 and seq
A;Reference number: I56203; MUID:94065175
A;Accession: I73012
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-683,'X',685-1133,'X',1135-1471,'X',1474-2489 <VIK1>
A;Cross-references: GB:L17418; NID:g305673; PIDN:AAB60695.1; PID:g451403
A;Accession: I56203
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-683,'X',685-894,'A',896-1033,1451-1471,'X',1474-2489 <VIK2>
A;Cross-references: GB:L17418; NID:g305673; PIDN:AAB60694.1; PID:g305680

R;Wong, W.W.; Cahill, J.M.; Kolson, M.D.; Kennedy, C.A.; Bonavida, B.I.; Morris, M.J.; Wi
J. Exp. Med. 169, 847-863, 1989
A;Title: Structure of the human CR1 gene. Molecular basis of the structural and quantita
A;Reference number: A47602; MUID:89176869
A;Accession: A47602
A;Molecule type: DNA
A;Residues: 1-41 <WON>
A;Cross-references: GB:X14855
R;Bourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b recep
type 1.
A;Reference number: S03291; MUID:89010527
A;Accession: S03291
A;Molecule type: mRNA
A;Residues: 26-584 <HOU>
A;Cross-references: EMBL:X14462; NID:g30197; PIDN:CAA32541.1; PID:g736249
A;Experimental source: clone CR1.4
R;Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A;Title: Identification of distinct C4b and C4b recognition sites in the human C3b/C4b r
A;Reference number: S03843; MUID:89035992
A;Accession: S03843
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-894,'A',896-1000,1451-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2
A;Cross-references: EMBL:Y00816; NID:g40185; PIDN:CAA68755.1; PID:g40186
R;Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A;Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domain
A;Reference number: A28507; MUID:87168191
A;Accession: A28507
A;Molecule type: mRNA
A;Residues: 954-1221,'EV',1224-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2327-2
A;Cross-references: GB:X05453; NID:g30196; PIDN:CAA28933.1; PID:g809019
R;Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A;Title: Identification of a partial cDNA clone for the human receptor for complement 3
A;Reference number: A94073; MUID:86067975
A;Accession: A24748
A;Molecule type: mRNA
A;Residues: 41-333,729-745,831-845 <WO2>
A;Cross-references: GB:M11569; NID:g180991; PIDN:AAA52297.1; PID:g180995; GB:M11617; NID
C;Genetics:
A;Gene: GDB:CR1; CD35
A;Cross-references: GDB:119080A; OMIM:120620
A;Map position: 1q42-1q42
A;Introns: 41/1; 207/1; 334/1; 1553/1; 2071/1; 2155/1; 2264/1; 2326/1; 2401/1; 2511/1; 2684/2;
; 1484/2; 1515/1; 1646/1; 1775/1; 1796/2; 1844/1; 1953/1; 1995/1; 2099/1
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-41/Domain: signal sequence #status predicted <SIG>
F;42-2489/Product: complement C3b/C4b receptor, membrane bound form #status predicted <M
F;42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MAT>
F;43-99/Domain: complement factor H repeat homology <FH01>
F;104-161/Domain: complement factor H repeat homology <FH02>
F;166-232/Domain: complement factor H repeat homology <FH03>
F;238-293/Domain: complement factor H repeat homology <FH04>
F;297-354/Domain: complement factor H repeat homology <FH05>
F;358-416/Domain: complement factor H repeat homology <FH06>
F;421-487/Domain: complement factor H repeat homology <FH07>
F;494-549/Domain: complement factor H repeat homology <FH08>
F;554-611/Domain: complement factor H repeat homology <FH09>
F;615-682/Domain: complement factor H repeat homology <FH10>
F;688-743/Domain: complement factor H repeat homology <FH11>
F;747-804/Domain: complement factor H repeat homology <FH12>
F;808-866/Domain: complement factor H repeat homology <FH13>
F;871-937/Domain: complement factor H repeat homology <FH14>
F;944-999/Domain: complement factor H repeat homology <FH15>
F;1004-1061/Domain: complement factor H repeat homology <FH16>
F;1066-1132/Domain: complement factor H repeat homology <FH17>
F;1138-1193/Domain: complement factor H repeat homology <FH18>
F;1197-1254/Domain: complement factor H repeat homology <FH19>
F;1258-1326/Domain: complement factor H repeat homology <FH20>
F;1321-1387/Domain: complement factor H repeat homology <FH21>
F;1393-1449/Domain: complement factor H repeat homology <FH22>
F;1454-1511/Domain: complement factor H repeat homology <FH23>
F;1516-1582/Domain: complement factor H repeat homology <FH24>
F;1588-1643/Domain: complement factor H repeat homology <FH25>
F;1647-1703/Domain: complement factor H repeat homology <FH26>
F;1708-1766/Domain: complement factor H repeat homology <FH27>
F;1771-1837/Domain: complement factor H repeat homology <FH28>
F;1846-1902/Domain: complement factor H repeat homology <FH29>
F;1907-1964/Domain: complement factor H repeat homology <FH30>
F;1969-2035/Domain: complement factor H repeat homology <FH31>
F;2041-2096/Domain: complement factor H repeat homology <FH32>
F;2100-2156/Domain: complement factor H repeat homology <FH33>
F;2161-2219/Domain: complement factor H repeat homology <FH34>
F;2224-2290/Domain: complement factor H repeat homology <FH35>
F;2298-2354/Domain: complement factor H repeat homology <FH36>
F;2359-2415/Domain: complement factor H repeat homology <FH37>

Query Match 29.3%; Score 220; DB 2; Length 2489;
Best Local Similarity 37.4%; Pred. No. 1.1e-12;
Matches 52; Conservative 16; Mismatches 55; Indels 16; Gaps 6;

Qy 2 IS[illegible]PPILN[illegible]PISYYS---TPIAV[illegible]TVIRYSC --SCTFRLICEKSLI[illegible]TIKDKV 52
Db 1967 IS[illegible]PFIISNG LEYSNNRTSFENGTVVTYQCHTGPDGEQLFELVGERSIYVTSKDGQ 2024

Qy 53 [illegible]WIKIAIKGIEYFNKYSSCPEFIVIGGYKIKGLE YEHCAVIIATKINPSMNGNKSV 111
Db 2025 VAVWSTFPPSCISTNK CTAPEVENAIPVPSNPSFPSLIEIVRFRCQPGFVMVGSHTV 2081

Qy 112 WCQANNMWGPTRLPTCVSV 130
Db 2082 QCQTNGRWGP-KLPHCSRV 2099

RESULT 6
I36936
complement receptor 1 - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I36936; I36937
R;Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J. Immunol. 153, 691-700, 1994
A;Title: Primary sequence of an alternatively spliced form of CR1. Candidate for the
A;Reference number: I36935; MUID:94292799
A;Accession: I36936
A;Status: preliminary; translated [illegible]; [illegible]MIM:DDBJ
A;Molecule type: mRNA
A;Residues: 1-2014 <BIR1>
A;Cross-references: GB:L24920; NID:g551564; PIDN:AAA51438.1; PID:g557729
A;Accession: I36937
A;Status: preliminary; translated [illegible]; [illegible]MIM:DDBJ
A;Molecule type: mRNA
A;Residues: 1-897,1751-2014 <BIR2>
A;Cross-references: GB:L24921; NID:g557726; PIDN:AAA51439.1; PID:g557727
C;Genetics:
A;Gene: CR1
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: duplication; glycoprotein; tandem repeat
F;18-74/Domain: complement factor H repeat homology <FH01>
F;79-136/Domain: complement factor H repeat homology <FH02>
F;141-207/Domain: complement factor H repeat homology <FH03>
F;213-268/Domain: complement factor H repeat homology <FH04>
F;272-328/Domain: complement factor H repeat homology <FH05>
F;333-391/Domain: complement factor H repeat homology <FH06>
F;396-462/Domain: complement factor H repeat homology <FH07>
F;470-526/Domain: complement factor H repeat homology <FH08>
F;529-586/Domain: complement factor H repeat homology <FH09>
F;722-778/Domain: complement factor H repeat homology <FH10>
F;1041-1107/Domain: complement factor H repeat homology <FH11>
F;1172-1228/Domain: complement factor H repeat homology <FH12>
F;1234-1291/Domain: complement factor H repeat homology <FH13>

```
F;1296-1362/Domain: complement factor H repeat homology <FH14>
F;1432-1489/Domain: complement factor H repeat homology <FH15>
F;1625-1681/Domain: complement factor H repeat homology <FH16>
F;1749-1815/Domain: complement factor H repeat homology <FH17>
F;1823-1879/Domain: complement factor H repeat homology <FH18>
F;1884-1940/Domain: complement factor H repeat homology <FH19>

  Query Match             29.0%;  Score 218;  DB 2;  Length 2014;
  Best Local Similarity   37.4%;  Pred. No. 1.4e-12;
  Matches   52;  Conservative   16;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPIAV   GTVIRYSC----- SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |    |||: | |          | |:||:|: | :||
Db    1492 ISCEPPTTISNG--DFYSNNRASFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1549

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
           | |  | ||:|   ||   |  | |    :: |:   :   : | | |:  | | |: :|
Db    1550 VGVWSSPPPRCISINK---CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 1606

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1607 QCQTNGRWGP-KLPHCSRV 1624


RESULT 7
WMVZSP
apolipoprotein H homolog precursor - vaccinia virus
N;Alternate names: 35K secretory protein; C3L protein; virokine
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999
C;Accession: A31005; B42504
R;Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A;Title: Vaccinia virus encodes a secretory polypeptide structurally related to compleme
A;Reference number: A31005; MUID:88318974
A;Accession: A31005
A;Molecule type: DNA
A;Residues: 1-263 <KOT>
A;Cross-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691
A;Experimental source: strain WR
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A;Title: Appendix to "The complete DNA sequence of vaccinia virus".
A;Reference number: A42501
A;Accession: B42504
A;Molecule type: DNA
A;Residues: 1-263 <GOE>
A;Cross-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345
A;Experimental source: strain Copenhagen
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A;Title: The complete DNA sequence of vaccinia virus
A;Reference number: A42501; MUID:91021027
A;Contents: annotation; possible protein coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: duplication; extracellular protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-263/Product: C4b binding protein homolog #status predicted <MAT>
F;21-81/Domain: complement factor H repeat homology <FH1>
F;86-143/Domain: complement factor H repeat homology <FH2>
F;148-201/Domain: complement factor H repeat homology <FH3>
F;206-261/Domain: complement factor H repeat homology <FH4>

  Query Match             28.6%;  Score 215;  DB 1;  Length 263;
  Best Local Similarity   35.9%;  Pred. No. 3e-13;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : | ||| | ||| : |      |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLC-----SGGEWSDP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   |    | : |:  | ||  : ::|: |  |   | |
Db     200 TCQIV----KCPHPTISNGYLSSGFKRSYSYNQNVEFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 8
C36838
complement contro. protein homolog - variola virus (strain India-1967)
N;Alternate names: D12L protein
C;Species: variola virus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 23-Mar-2001
C;Accession: C36838
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: C36838
A;Molecule type: DNA
A;Residues: 1-263 <BLI>
A;Cross-references: GB:X69198; NID:g456756; PIDN:CAA48954.1; PID:g297195
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
F;21-81/Domain: complement factor H repeat homology <FH1>
F;86-143/Domain: complement factor H repeat homology <FH2>
F;148-201/Domain: complement factor H repeat homology <FH3>
F;206-261/Domain: complement factor H repeat homology <FH4>

  Query Match             28.0%;  Score 210;  DB 1;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 9.2e-13;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : | ||| | ||| : |      |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLC-----SGGEWSNP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       || | :  ||   |    | : |:| | ||  : ::|: |  |   | |
Db     200 TCQIV----KCPHPTILNGYLSSGFKRSYSYNQNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 9
T28450
hypothetical protein D15L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28450
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Ac
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox vir
A;Reference number: Z20488; MUID:94088747
A;Accession: T28450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-263 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60766.1; PID:g438955
A;Experimental source: strain "Bangladesh-1975"
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom

  Query Match             28.0%;  Score 210;  DB 2;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 9.2e-13;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : | ||| | ||| : |      |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLC-----SGGEWSNP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
```

```
                 ||:       || | :  :    |    | : |:| | ||  : ::|: |  |   | |
Db     200 DQQIV    KCPYPTILNGYISSGFKRSYSYNQNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           :| ||
Db     256 P ELPKCV 262


RESULT 10
B72152
B18L protein - variola minor virus (strain Garcia 1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: B72152
R;Shchelkunov, S.N.; Totmenin, A.V.; Safronov, P.F.; Mikheev, M.V.; Gut
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
A;Reference number: A72150
A;Accession: B72152
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54612.1; PID:g5830573
A;Experimental source: strain Garcia 1966
C;Genetics:
A;Gene: B18L
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo

  Query Match             27.7%;  Score 208;  DB 2;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 1.4e-12;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 LSCGSPPTILNGRISYYSTPI-AVGTVIKYSCSGTFKLIGEKSIICTTKDKVDGTWDKPAP 60
           : |  |: | ||| :  |:      |:|: ||||  : |||   :|||        | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVIC      SGGEWSNP-P 199

Qy      61 PKCEYFNFYSSCPPPIVPGGYKIRGSTP-YRPGDSVTFACKTNFSMNGNKSVWCQANNMW 119
           |:       || | :  ||    |    | : |:| | ||  : ::|: |  |   | |
Db     200 DQQIV    KCPYPTILNGYISSGFKRSYSYNQNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     120 GPTRLPTCV 128
           |  || :|
Db     256 P ELPKCV 262


RESULT 11
A56740
sperm egg recognition protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 20-Aug-1999
C;Accession: A56740
R;Bookbinder, L.H.; Cheng, A.; Bleil, J.D.
Science 269, 86-89, 1995
A;Title: Tissue- and species-specific expression of sp56, a mouse sperm fertilization pr
A;Reference number: A56740; MUID:95327956
A;Accession: A56740
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-579 <BOO>
A;Cross-references: GB:U17108; NID:g897562; PIDN:AAC52208.1; PID:g897563
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: fertilization; homomultimer; peripheral membrane protein; sperm
F;1-42/Domain: signal sequence #status predicted <SIG>
F;44-91/Domain: complement factor H repeat homology <FH1>
F;96-153/Domain: complement factor H repeat homology <FH2>
F;158-218/Domain: complement factor H repeat homology <FH3>
F;223-278/Domain: complement factor H repeat homology <FH4>
F;283-345/Domain: complement factor H repeat homology <FH5>
F;349-411/Domain: complement factor H repeat homology <FH6>
F;456-509/Domain: complement factor H repeat homology <FH7>

  Query Match             27.1%;  Score 203.5;  DB 2;  Length 579;
  Best Local Similarity   33.1%;  Pred. No. 9.1e-12;
  Matches   42;  Conservative   22;  Mismatches   56;  Indels    7;  Gaps    4;

Qy       3 SCGSPPTILNGRISYYSTPI-AVGTVIKYSCSGTFKLIGEKSIICTTKDKVDGTWDKPAP 61
           :|  |||| ||: |    :   |:|: |:|   || |:|   |:::    :: | |    |
Db     152 TCELPFTFENGKHESGRQVDDLYTEGSVVIYNCDEGFELLGNASIFCTVVNKTVGVWRPHPP 211

Qy      62 KCEYFNFYSSCPPPIVPGGYKIRGSTP-YRPGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:       |  | :| ||    |    | : |::    || |  :  : :|||  |
Db     212 ACQKLE--VCHKPGTPRGYLAVGFPGSYAYRSAITFRCKKGFELRGSSVIRCEANCEWP 272

Qy     121 PTRLPTCV 127
           |: :|||
Db     273 PS-IPTC 278


RESULT 12
S53711
C4BP alpha chain precursor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S53711
R;de Frutos, P.G.; Dahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A;Title: cDNA structure of rabbit C4b-binding protein alpha-chain: Preserved sequence
A;Reference number: S53711; MUID:95226458
A;Accession: S53711
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-597 <DEF>
A;Cross-references: EMBL:Z35490
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;50-107/Domain: complement factor H repeat homology <FH1>
F;112-169/Domain: complement factor H repeat homology <FH2>
F;174-234/Domain: complement factor H repeat homology <FH3>
F;239-294/Domain: complement factor H repeat homology <FH4>
F;299-360/Domain: complement factor H repeat homology <FH5>
F;364-422/Domain: complement factor H repeat homology <FH6>
F;426-480/Domain: complement factor H repeat homology <FH7>
F;484-538/Domain: complement factor H repeat homology <FH8>

  Query Match             27.0%;  Score 202.5;  DB 1;  Length 597;
  Best Local Similarity   33.8%;  Pred. No. 1.2e-11;
  Matches   44;  Conservative   21;  Mismatches   58;  Indels    7;  Gaps    4;

Qy       2 LSCGSPPTILNGRISYYSTPI-AVGTVIKYSCSGTFKLIGEKSIICTTKDKVDGTWDKPAP 60
           : |  |: | ||| :  |:      |:|: ||||  : |||   :|||        | |  | |
Db     172 VKCEPPFNFENGKHESGRQVDDLYTEGSVVIYSCNSGYSLIGNASIFCTVVNKTVGVWRPSPP 231

Qy      61 PKCEYFNFYSSCPPPIVPGGYKIRGSTP-YRPGDSVTFACKTNFSMNGNKSVWCQANNMW 119
           |:       |  | :| ||    |    | : |::    || |  :  : :|||  |
Db     232 PVC- --KEITCSPPNVPHGKIISGFGPIYNYKDSIMYICIDCFVLRGSSLIRCELDSKW 287

Qy     120 GPTRLPTCVS 129
           |:      |
Db     288 NPSP-PVCES 296


RESULT 13
A34924
complement C3b/C4b receptor-like protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 27-Jul-1990 #sequence_revision 11-Apr-1997 #text_change 03-Dec-1999
C;Accession: A34924; S03292
R;Bourcade, D.; Miesner, D.R.; Bee, C.; Zoldes, W.; Atkinson, J.P.
J. Biol. Chem. 265, 974-980, 1990
A;Title: Duplication and divergence of the amino terminal coding region of the comple
A;Reference number: A34924; MUID:90110163
A;Accession: A34924
A;Molecule type: DNA
A;Residues: 1-479 <BEU>
A;Cross-references: GB:J05195
R;Bourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
```

```
A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
 type 1.
A;Reference number: S03291; MUID:89010527
A;Accession: S03292
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 'FPFFAFR',452-482 <H02>
A;Cross-references: EMBL:X14360
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1988
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;43-99/Domain: complement factor H repeat homology <FH01>
F;104-161/Domain: complement factor H repeat homology <FH02>
F;166-232/Domain: complement factor H repeat homology <FH03>
F;238-293/Domain: complement factor H repeat homology <FH04>
F;297-353/Domain: complement factor H repeat homology <FH05>
F;358-416/Domain: complement factor H repeat homology <FH06>
F;421-478/Domain: complement factor H repeat homology <FH07>

  Query Match          26.8%;  Score 291;  DB 2;  Length 482;
  Best Local Similarity   36.5%;  Pred. No. 1.3e-11;
  Matches   50;  Conservative   15;  Mismatches   60;  Indels   12;  Gaps    5;

Qy       2 ISPGSPPPTLNGRTSVYSTP-IAVGTVTRYSCS-----GTFRLIGEKSLICTTKDKVDG 54
           | || || | ||   :  |      |:|: | |:       | |:|| |: | :||   |
Db     164 IICGLPPTIANGDFTSISREYFHYGSVVTYHCNLGSRGKKVFELVGEPSIYCTSKDDQVG 223

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
           |  |||:|   ||   | | |  |  :  : :   - | | |:  | | |   | |
Db     224 IWSGPAPQCIIPNK--SSPENVENGILVSDNRSLFSLNEVVEFRCQPGFHMKGPSHVKC 280

Qy     114 QANNMWGPTRLPTCVSV 130
           || | | |  ||:|  |
Db     281 QALNKWEP-ELPSCSRV 296


RESULT 14
NBHUC4
C4b-binding protein alpha chain precursor - human
N;Alternate names: C4BP; proline rich protein
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A33568; S02372; A90326; A24182; A93134; S29492; A31785; I52244; A03210
R;Matsuguchi, T.; Okamura, S.; Aso, T.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 165, 138-144, 1989
A;Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity o
A;Reference number: A33568; MUID:90073699
A;Accession: A33568
A;Molecule type: mRNA
A;Residues: 1-597 <MA1>
A;Cross-references: GB:M31452; NID:g190501; PIDN:AAA36507.1; PID:g190502
A;Note: the authors translated the codon GGA for residue 492 as Glu
R;Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.
FEBS Lett. 232, 328-332, 1988
A;Title: Derivation of the sequence of the signal peptide in human C4b-binding protein a
A;Reference number: S02372; MUID:88242821
A;Accession: S02372
A;Molecule type: mRNA
A;Residues: 17-81 <LI2>
A;Cross-references: EMBL:X07853
A;Note: although the sequence determined extends to residue 9 above, these authors assig
R;Chung, L.P.; Bentley, D.R.; Reid, K.B.M.
Biochem. J. 230, 133-141, 1985
A;Title: Molecular cloning and characterization of the cDNA coding for C4b-binding prote
A;Reference number: A90326; MUID:86025405
A;Accession: A90326
A;Molecule type: mRNA
A;Residues: 80-597 <CH2>
A;Cross-references: GB:X02865; NID:g29564; PIDN:CAA26617.1; PID:g29565
A;Note: 92 Thr and 457 His were also found
R;Lintin, S.J.; Reid, K.B.M.
FEBS Lett. 204, 77-81, 1986
A;Title: Studies on the structure of the human C4b-binding protein gene.
A;Reference number: A24182; MUID:86301119
A;Accession: A24182
A;Molecule type: DNA
A;Residues: 203-288 <LIN>
A;Cross-references: EMBL:X04284; EMBL:X04296
R;Rodriguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.
J. Exp. Med. 173, 1073-1082, 1991
A;Title: Structure of the gene coding for the alpha polypeptide chain of the human co
A;Reference number: A43023; MUID:91217619
A;Contents: annotation; exon intron boundaries
R;Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol. Immunol. 22, 427-435, 1985
A;Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequenc
ced by cyanogen bromide treatment.
A;Reference number: A93134; MUID:85296005
A;Accession: A93134
A;Molecule type: protein
A;Residues: 49-81 <CH1>
A;Note: this paper reports amino terminal sequences of the intact protein and of a cu
R;Dessing, M.; Kanters, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B
FEBS Lett. 317, 228-232, 1993
A;Title: The region Ser(333)-Arg(356) of the alpha chain of human C4b-binding protein
A;Reference number: S29492; MUID:93146164
A;Accession: S29492
A;Status: preliminary
A;Molecule type: protein
A;Residues: 381-404 <HES>
R;Suzuki, K.; Nishioka, J.
J. Biol. Chem. 263, 17034-17039, 1988
A;Title: Binding site for vitamin K-dependent protein S on complement C4b-binding pro
A;Reference number: A31785; MUID:89034204
A;Accession: A31785
A;Molecule type: protein
A;Residues: 495-505,'X',507-510,'X',512-515 <SUZ>
A;Note: this peptide appears to bind protein S
R;Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983
A;Title: Visualization of human C4b-binding protein and its complexes with vitamin K-
A;Reference number: A93950; MUID:83221615
A;Contents: annotation; electron microscopy, three-dimensional structure; ligand bind
R;Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 174, 222-227, 1991
A;Title: Genomic organization of the alpha chain of the human C4b-binding protein gen
A;Reference number: I52244; MUID:91113199
A;Accession: I52244
A;Status: translation not shown, translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-597 <ASO>
A;Cross-references: GB:M62486; NID:g190498; PIDN:AAA36506.1; PID:g190500
C;Comment: C4BP controls the classical pathway of complement activation. It binds as
 the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement
C;Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment
isulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. Amino
actors V and VIII.
C;Comment: The molecule has a central body support ng seven tentacles (alpha chains).
C;Genetics:
A;Gene: GDB:C4BPA
A;Cross-references: GDB:120568; OMIM:120830
A;Map position: 1q32-1q32
A;Introns: 48/1; 130/1; 143/2; 172/1; 230/1; 247/1; 362/1; 425/1; 481/1; 543/3
C;Complex: octamer of seven alpha chains and one beta chain
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein;
F;1-48/Domain: signal sequence #status predicted <SIG>
F;49-597/Product: C4b binding protein alpha chain #status predicted <MAT>
F;50-108/Domain: complement factor H repeat homology <FH1>
F;113-170/Domain: complement factor H repeat homology <FH2>
F;175-234/Domain: complement factor H repeat homology <FH3>
F;238-294/Domain: complement factor H repeat homology <FH4>
F;299-360/Domain: complement factor H repeat homology <FH5>
F;364-422/Domain: complement factor H repeat homology <FH6>
F;381-404/Region: complement C4b binding #status predicted
F;426-480/Domain: complement factor H repeat homology <FH7>
F;484-539/Domain: complement factor H repeat homology <FH8>
```

F;221/Sol: 528/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 26.0%; Score 199; DB 1; Length 597;
Best Local Similarity 32.8%; Pred. No. 2.5e-11;
Matches 44; Conservative 28; Mismatches 60; Indels 6; Gaps 6;

Qy 2 ISCGSPPPTLNGRISYYSI-TPIAVGTVIRYSCSGTFRLIGEKSLLCTTKDKVIWDKPAP 61
Db 170 VKCKPEPPDIRNGRHSGEEN YAYGESVIYSCDPRFSLIGHASISCIVENETIGVWRPSPP 232

Qy 62 KCEYENKYSSCPEPIVPGGYKIRGSIP YRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
Db 233 ICEKI ICRKPDVSHGCMVSGFGVDLYNYKIDIVFKCQKGFVLRGSSVIHCDADSKWN 288

Qy 121 PTELPTC 127
Db 289 PSP PAC 294

RESULT 15
I46001
C4b binding protein alpha chain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: I46001; S43190
R;Hillarp, A.; Thern, A.; Dahlback, B.
J. Immunol. 153, 4190-4199, 1994
A;Title: Bovine C4b-binding protein: Molecular cloning of the alpha- and beta-chains pro
A;Reference number: I46001; MUID:95015909
A;Accession: I46001
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-610 <HIL>
A;Cross-references: EMBL:Z31693; NID:g469117; PIDN:CAA84498.1; PID:g469118
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
F;50-107/Domain: complement factor H repeat homology <FH1>
F;111-169/Domain: complement factor H repeat homology <FH2>
F;171-234/Domain: complement factor H repeat homology <FH3>
F;236-294/Domain: complement factor H repeat homology <FH4>
F;296-362/Domain: complement factor H repeat homology <FH5>
F;365-425/Domain: complement factor H repeat homology <FH6>
F;426-484/Domain: complement factor H repeat homology <FH7>
F;487-541/Domain: complement factor H repeat homology <FH8>

Query Match 26.0%; Score 197.5; DB 1; Length 610;
Best Local Similarity 32.8%; Pred. No. 3.6e-11;
Matches 44; Conservative 21; Mismatches 52; Indels 15; Gaps 5;

Qy 4 CGSPPPTLNGRISYYSITPIAVGTVIRYSCSGTFRLIGEKSLLCTTKDKVIWDKP 58
Db 174 CEPPPIINGRHSGEDFEYAYGESVIYSCDRFFSLIGHASISCIVENETIGVWSP 229

Qy 59 PAPKCEYENKYSSCPEPIVPGGYKIRGSIP YRHGDSVTFACKTNFSMNGNKSVWCQANN 117
Db 230 SPPSC KKVIVQPVVPTYCEIISGFGPIYEYQQSIVYACNEGFELEGDSILHCEADN 285

Qy 118 MWGPTRLPTC 127
Db 286 SWNPPP PPC 294

RESULT 16
JC2054
complement regulatory protein, 512 antigen precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C;Accession: JC2054; PC2027
R;Sakakibara, T.; Seno, H.; Dohi, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.
Biochem. Biophys. Res. Commun. 198, 819-826, 1994
A;Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.
A;Reference number: JC2054; MUID:94161746
A;Accession: JC2054
A;Molecule type: mRNA
A;Residues: 1-497 <SAK>
A;Cross-references: GB:D42114; NID:g577641; PIDN:BAA07688.1; PID:g987688
A;Accession: PC2027
A;Molecule type: protein
A;Residues: 39-51 <SA2>
C;Comment: This protein plays a critical role in protection against complement mediat
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: glycoprotein; transmembrane protein
F;1-38/Domain: signal sequence #status predicted <SIG>
F;37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>
F;38-94/Domain: complement factor H repeat homology <FH1>
F;99-156/Domain: complement factor H repeat homology <FH2>
F;161-227/Domain: complement factor H repeat homology <FH3>
F;233-288/Domain: complement factor H repeat homology <FH4>
F;294-352/Domain: complement factor H repeat homology <FH5>
F;357-413/Domain: complement factor H repeat homology <FH6>
F;421-444/Domain: transmembrane #status predicted <TMM>
F;247,331,346,450,482,484/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 24.5%; Score 184; DB 2; Length 497;
Best Local Similarity 33.1%; Pred. No. 5.7e-10;
Matches 46; Conservative 19; Mismatches 58; Indels 16; Gaps 6;

Qy 2 ISCGSPPPTLNGRISYYS - TPIAVGTVIRYSCSG GTFRLIGEKSLLCTTKDKV 52
Db 159 IPCEIPPSIPNG- DPFSPNREDFHYGMVVTYQCNTPARGKKLFNLVGEPSIHCTSIDGQ 216

Qy 53 DGIWTPEPAEPCEYENKYSSCPEPIVPGGYKI PYRHGDSVTFACKTNFSMNGNKSV 111
Db 217 VGVWSGTALPICEVNK -CPHVLNAVIVSKNK--TKYGVVEYKCKPGMMFDGSSV 273

Qy 112 WCQANNMWGPTRLPTCVSV 130
Db 274 YCPSLNPWEP-QLPSCFKV 291

RESULT 17
S57953
C4BP protein alpha chain precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C;Accession: S57953
R;Hillarp, A.; Thern, A.; Dahlback, B.
submitted to the EMBL Data Library, July 1995
A;Description: Molecular cloning of rat C4b binding protein alpha- and beta-chains: s
A;Reference number: S57953
A;Accession: S57953
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-558 <HIL>
A;Cross-references: EMBL:Z50651; NID:g899379; PIDN:CAA90491.1; PID:g899380
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
F;15-72/Domain: complement factor H repeat homology <FH1>
F;77-134/Domain: complement factor H repeat homology <FH2>
F;139-199/Domain: complement factor H repeat homology <FH3>
F;204-258/Domain: complement factor H repeat homology <FH4>
F;263-324/Domain: complement factor H repeat homology <FH5>
F;328-386/Domain: complement factor H repeat homology <FH6>
F;390-443/Domain: complement factor H repeat homology <FH7>
F;447-501/Domain: complement factor H repeat homology <FH8>

Query Match 24.2%; Score 182; DB 2; Length 558;
Best Local Similarity 29.9%; Pred. No. 1e-09;
Matches 41; Conservative 18; Mismatches 48; Indels 30; Gaps 6;

Qy 4 CGSPPPTLNGK--------ISYYSITPIAVGTVIRYSCSGTFRLIGEKSLLCTTKDKVIWGK 55
Db 139 CGMPPDISNGKHNGREEEPFTYKSS-- VIYKCHEGFTLIGNASIICIVVNKTVGV 191

Qy 56 WDKPAPKCEYENKYSSCPEP--- IVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKS 110
Db 192 WSPSPPICERI ---ICPWPEVLHGTINSCPK- HIYKYKCRVPFVCQRGIVLRGSGV 243

```
Qy      111 VWCQANNMWGPTRLPTC 127
            : |:|:  | |  :| |
Db      244 IHCEADGSWSP--VPVC 258


RESULT 18
A43519
complement receptor CR1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-Oct-1992 #sequence_revision 30-Jan-1993 #text_change 17-Nov-2000
C;Accession: A43519; A30550
R;Paul, M.S.; Aeqerter, M.; Cepek, K.; Miller, M.D.; Weis, J.H.
J. Immunol. 144, 1988-1996, 1990
A;Title: The murine complement receptor qene family. The qenomic and transcriptional com
A;Reference number: A43519; MUID:90171600
A;Accession: A43519
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-440 <PAU>
A;Cross-references: GB:M34164
A;Note: the authors translated the codon GGC for residue 21 as Ala, and CAG for residue
R;Paul, M.S.; Aeqerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.
J. Immunol. 142, 582-589, 1989
A;Title: The murine complement receptor qene family. Analysis of mCRY qene products and
A;Reference number: A30550; MUID:89093944
A;Accession: A30550
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-20,'A',22-120,'E',122-473 <PA2>
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;42-98/Domain: complement factor H repeat homology <FH1>
F;103-160/Domain: complement factor H repeat homology <FH2>
F;165-231/Domain: complement factor H repeat homology <FH3>
F;237-293/Domain: complement factor H repeat homology <FH4>
F;299-355/Domain: complement factor H repeat homology <FH5>

  Query Match             24.0%;  Score 180.5;  DB 2;  Length 440;
  Best Local Similarity   34.6%;  Pred. No. 1.1e-09;
  Matches   47;  Conservative   14;  Mismatches   60;  Indels   15;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYST--PIAVGTVIRYSCS------GTFRLIGEKSLLCITKDKV 52
           |  | || | ||   :::|:       | |: | |:        | |:|| || | : |
Db     163 IPCEIPPGIPNG--DFFSSTREDFHYGMVVTYRCNTDARGKALFNLVGEPSLYCTSNDGE 220

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKI-RGSTPYRHGDSVTFACKTNFSMNGNKSV 111
           |  | |:|   || : | | |      :   : :  | | | |   | | |  ||
Db     221 IGVWSGPPPQCIELNKCT--PPPVVENAVMLSENRSLFSLRDIVEFRCHPGFIMKGASSV 278

Qy     112 WCQANNMWGPTRLPTC 127
           ||: | | |  |:|
Db     279 HCQSLNKWEP ELPSC 293


RESULT 19
NBHUH
complement factor H precursor, long splice form [validated] - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec 2000
C;Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R;Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A;Title: The complete amino acid sequence of human complement factor H.
A;Reference number: S00254; MUID:88134059
A;Accession: S00254
A;Molecule type: mRNA
A;Residues: 1-1241 <RIP>
A;Cross-references: EMBL:Y00716; NID:q31964; PIDN:CAA68704.1; PID:q31965
A;Note: 402 Tyr was also found
A;Note: parts of this sequence, including the amino and carboxyl ends of the mature prot
R;Estaller, C.; Schwaeble, W.; Dierich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A;Title: Human complement factor H: two factor H proteins are derived from alternatively
A;Reference number: A60238; MUID:91184292
A;Accession: A60238
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-56;.177-1231 <EST>
A;Note: only portions of this 4.3 kilobase mRNA were sequenced
R;Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, E.; Harris, T.J.R.; Sim, R.B.
Biosci. Rep. 7, 201-207, 1987
A;Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl
A;Reference number: A54726; MUID:88025472
A;Accession: A54726
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'DFRN',579-1231 <DAY>
A;Cross-references: GB:M17517; NID:q180497; PIDN:AAA52016.1; PID:q180498
A;Note: parts of this sequence were determined by protein sequencing
R;Ripoche, J.; Day, A.J.; Willis, A.C.; Bent, K.T.; Campbell, R.D.; Sim, R.B.
Biosci. Rep. 6, 65-72, 1986
A;Title: Partial characterization of human complement factor H by protein and cDNA se
A;Reference number: A61565; MUID:86188124
A;Accession: A61565
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'MITGRNHLNAKI',1050-1057,'?',1059-1112 <RIP2>
R;Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A;Title: Purification and structural studies on the complement-system control protein
A;Reference number: A26505; MUID:83048213
A;Accession: A26505
A;Molecule type: protein
A;Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>
R;Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P
Biochemistry 31, 3626-3634, 1992
A;Title: Solution structure of the fifth repeat of factor H: A second example of the
A;Reference number: A44551; MUID:92232649
A;Contents: annotation, NMR structure determination, residues 264-292
R;Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A;Title: Three dimensional structure of a complement control protein module in soluti
A;Reference number: A49224; MUID:91278097
A;Contents: annotation; NMR structure determination, residues 927-985
R;Estaller, C.; Koistinen, V.; Schwaeble, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A;Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no

A;Reference number: I56100; MUID:91201892
A;Accession: I72654
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1047-1231 <RES>
A;Cross-references: GB:M65294; NID:q183768; PIDN:AAA35948.1; PID:q183767
R;Carron, J.A.; Bates, R.C.; Smith, A.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Born
Biochim. Biophys. Acta 1289, 305-311, 1996
A;Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A;Reference number: S66298; MUID:96205165
A;Accession: S66298
A;Status: preliminary
A;Molecule type: protein
A;Residues: 411-419;574-578,580-582 <CAR>
C;Comment: Factor H has also been found bound to cell membranes in an unknown manner.
C;Comment: Alternative transcipts of 4.3, 1.8, and 1.4 kilobases are expressed in liv
C;Genetics: <HF1>
A;Gene: GDB:HF1; HF
A;Cross-references: GDB:120041; OMIM:134370
A;Map position: q32-1q32
C;Genetics: <HF2>
A;Gene: GDB:HF2; HF
A;Cross-references: GDB:129095
A;Map position: 1q32-1q32
A;Note: the correspondence between the two loci and the sequences indicated is unclea
C;Function:
A;Description: a cofactor in the inactivation of C3b by serine proteinase I; also the
he alternative complement pathway
A;Pathway: complement alternate pathway
```

C;Superfamily: complement factor H; complement factor H repeat homology
C;Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1229/Product: complement factor H #status experimental <MPT>
F;19-449/Product: complement factor H, heat splice form #status experimental <MAT>
F;21-80/Domain: complement factor H repeat homology <FH01>
F;85-141/Domain: complement factor H repeat homology <FH02>
F;144-205/Domain: complement factor H repeat homology <FH03>
F;210-262/Domain: complement factor H repeat homology <FH04>
F;246-248/Region: cell attachment (R G D) motif
F;267-320/Domain: complement factor H repeat homology <FH05>
F;325-385/Domain: complement factor H repeat homology <FH06>
F;389-442/Domain: complement factor H repeat homology <FH07>
F;448-505/Domain: complement factor H repeat homology <FH08>
F;564-624/Domain: complement factor H repeat homology <FH09>
F;627-684/Domain: complement factor H repeat homology <FH10>
F;686-744/Domain: complement factor H repeat homology <FH11>
F;747-803/Domain: complement factor H repeat homology <FH12>
F;808-864/Domain: complement factor H repeat homology <FH13>
F;870-926/Domain: complement factor H repeat homology <FH14>
F;931-984/Domain: complement factor H repeat homology <FH15>
F;989-1044/Domain: complement factor H repeat homology <FH16>
F;1046-1102/Domain: complement factor H repeat homology <FH17>
F;1109-1163/Domain: complement factor H repeat homology <FH18>
F;1167-1228/Domain: complement factor H repeat homology <FH19>
F;41-66,82-89,129,134,141,146,192,198,205,210,231,237,262,267,309,294-320,325-374,85
1,804-811,853,859,864,870,915,901,926,931,923,959,984,989-1032,1018,1043,1048,1091,1077
F;217/Binding site: carbohydrate (Asn) (covalent) #status absent
F;529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;218,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.0%; Score 180.5; DB 1; Length 1231;
Best Local Similarity 34.8%; Pred. No. 3.4e-09;
Matches 47; Conservative 24; Mismatches 50; Indels 15; Gaps 8;

Qy 8 SCGSPPPELNGRISYYSITPTAVC 31
Db 629 SCGPPPELLNGNVK EKTKEEYGHSEVVEYYCNPRFLMKGPNKIQC---- VDGEW TTL 681

Qy 61 PKCEYFNKYSSCPS PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 119
Db 682 PVC IVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCSESFTMIGHRSITC-THGVW 738

Qy 120 GPTELPTCVSVEPLE 134
Db 739 TQLPQCVAIDKLK 751

RESULT 20
NBMSC4
C4b binding protein alpha chain precursor - mouse
N Alternate names: C4BP; proline rich protein
C;Species: Mus musculus (house mouse)
C Date: 30-Jun-1994 #sequence_revision 30-Jun-1993 #text_change 18-Jul-1997
C Accession: A27117
R Kristensen, T.; Ogata, R.T.; Chung, L.P.; Reid, K.B.M.; Tack, B.F.
Biochemistry 26, 4668-4674, 1987
A Title: cDNA structure of murine C4b binding protein, a regulatory component of the ser
A Reference number: A27117; MUID:88024997
A Accession: A27117
A Molecule type: mRNA
A Residues: 1-469 <KRI>
A Note: the authors translated the codon GCT for residue 25 as Val
C;Comment: C4BP controls the classical pathway of complement activation. It binds as a c
the degradation of the C4b*2a complex (C3 convertase) by dissociating the complement fr
C;Comment: In mouse, C4BP is a multimeric protein of noncovalently associated chains.
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla
F;1-56/Domain: signal sequence #status predicted <SIG>
F;57-469/Product: C4b binding protein alpha chain #status predicted <MAT>
F;58-115/Domain: complement factor H repeat homology <FH1>
F;120-176/Domain: complement factor H repeat homology <FH2>
F;181-240/Domain: complement factor H repeat homology <FH3>
F;245-299/Domain: complement factor H repeat homology <FH4>
F;303-355/Domain: complement factor H repeat homology <FH5>
F;359-413/Domain: complement factor H repeat homology <FH6>
F;74,227,271,342,366,381,426/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 23.8%; Score 178.5; DB 1; Length 469;
Best Local Similarity 33.3%; Pred. No. 1.8e-09;
Matches 45; Conservative 13; Mismatches 54; Indels 24; Gaps 7;

Qy 2 ISCSPPPELNGRIS--- YYSITPTAVCIVTRYSCSGTERLIGEKSLLCITKDKVDCTWD 57
Db 173 VRCGTPPDLLNGKHSGRDFEVGY EYKHG- ESVTYKCDPGLVGSDVIRTCQEDGRSAWSA 231

Qy 58 PKAPFCEYFNKYSSCTEI IVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 117
Db 232 SSPPTCEFI----PCSQPNILHGVIVSGYR ATYTHMDSVEFACTLNCTVLSGHHVIE 286

Qy 118 CQANNMWGPTRLPTC 132
Db 287 CQQNGNW SSSLPTC 299

RESULT 21
NBMSH
complement factor H precursor - mouse
N;Alternate names: protein beta-1-H
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999
C;Accession: A26154; I49711; I49728
R;Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A;Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt
A;Reference number: A26154; MUID:86233453
A;Accession: A26154
A;Molecule type: mRNA
A;Residues: 1-1234 <KRI>
A;Cross-references: GB:M12660; NID:g194724; PIDN:AAA37759.1; PID:g487181
R;Natsuume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A;Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
A;Reference number: I49711; MUID:90111033
A;Accession: I49711
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-18 <RES>
A;Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729
R;Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.
Biochemistry 28, 9891-9897, 1989
A;Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN gamma
A;Reference number: I49728; MUID:90148935
A;Accession: I49728
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-19 <RE2>
A;Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926
C;Comment: Two codominant alleles of factor H are present in mice.
C;Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot
(C5 convertase) in the alternative complement pathway.
C;Genetics:
A;Map position: 1
C;Superfamily: complement factor H; complement factor H repeat homology
C;Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1234/Product: complement factor H #status predicted <MPT>
F;21-80/Domain: complement factor H repeat homology <FH01>
F;85-141/Domain: complement factor H repeat homology <FH02>
F;146-205/Domain: complement factor H repeat homology <FH03>
F;210-262/Domain: complement factor H repeat homology <FH04>
F;246-248/Region: cell attachment (R-G-D) motif
F;267-320/Domain: complement factor H repeat homology <FH05>
F;325-385/Domain: complement factor H repeat homology <FH06>
F;389-442/Domain: complement factor H repeat homology <FH07>

```
F;448-505/Domain: complement factor H repeat homology <FH08>
F;509-564/Domain: complement factor H repeat homology <FH09>
F;569-622/Domain: complement factor H repeat homology <FH10>
F;629-683/Domain: complement factor H repeat homology <FH11>
F;690-743/Domain: complement factor H repeat homology <FH12>
F;752-802/Domain: complement factor H repeat homology <FH13>
F;808-861/Domain: complement factor H repeat homology <FH14>
F;867-941/Domain: complement factor H repeat homology <FH15>
F;936-989/Domain: complement factor H repeat homology <FH16>
F;994-1048/Domain: complement factor H repeat homology <FH17>
F;1053-1107/Domain: complement factor H repeat homology <FH18>
F;1114-1168/Domain: complement factor H repeat homology <FH19>
F;1172-1230/Domain: complement factor H repeat homology <FH20>
F;21-66,52-80,85-129,114-141,146-192,178-295,210-251,237-262,267-309,294-320,325-374,357-
0-802-808-850,836-861,867-920,906-941,936-978,964-989,994-1037,1023-1048,1053-1096,1082-
F;676-721,773-801,1040,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

  Query Match             23.8%;  Score 178.5;  DB 1;  Length 1234;
  Best Local Similarity   37.0%;  Pred. No. 5.1e-09;
  Matches   51;  Conservative   14;  Mismatches   52;  Indels   21;  Gaps    8;

Qy       3 SCGSPPPILNGRIS-----YYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWD 57
           || |  |||| |:       ||     | |::| |   | | |   : |     ||| |
Db     628 SCAPPLEILNGEINGAKKVEYSH---GEVVKYDCKPRFLLKGPNKIQC-----VDGNW- 677

Qy      58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQAN 116
             | |    :  :| : | :  |       || |||| | |: ||:| |: || |  :
Db     678 TTLPVC-IEEERTCGDIPELEHGSAKCSVPPYHHGDSVEFICEENFTMIGHGSVSC-IS 734

Qy     117 NMWGPTRLPTCVSVFPLE 134
           |  |:|| ||:   ||
Db     735 GKW-TQLPKCVATDQLE 750


RESULT 22
D35069
complement factor H-related protein 9C4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20 Aug-1999
C;Accession: D35069; A35070; B35070; C35070; D35070; E35069; E35070; F35069; F35070; G35
R;Vik, D.P.; Munoz-Canoves, P.; Koziner, B.; Martin, E.C.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 4193-4201, 1990
A;Title: Identification and sequence analysis of four complement factor H-related transc
A;Reference number: A35070; MUID:90153969
A;Accession: D35069
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-808 <VIK>
A;Cross-references: GB:M29009; NID:g192563; PIDN:AAA37416.1; PID:g309166; GB:J05259
A;Note: translation of the nucleotide sequence is not complete
C;Superfamily: complement factor H; complement factor H repeat homology
F;28-88/Domain: complement factor H repeat homology <FH01>
F;92-145/Domain: complement factor H repeat homology <FH02>
F;151-208/Domain: complement factor H repeat homology <FH03>
F;212-267/Domain: complement factor H repeat homology <FH04>
F;272-325/Domain: complement factor H repeat homology <FH05>
F;332-386/Domain: complement factor H repeat homology <FH06>
F;393-446/Domain: complement factor H repeat homology <FH07>
F;455-505/Domain: complement factor H repeat homology <FH08>
F;511-564/Domain: complement factor H repeat homology <FH09>
F;569-622/Domain: complement factor H repeat homology <FH10>
F;627-681/Domain: complement factor H repeat homology <FH11>
F;688-742/Domain: complement factor H repeat homology <FH12>
F;746-807/Domain: complement factor H repeat homology <FH13>

  Query Match             22.0%;  Score 165.5;  DB 2;  Length 808;
  Best Local Similarity   46.7%;  Pred. No. 5.8e-08;
  Matches   51;  Conservative    9;  Mismatches   56;  Indels   23;  Gaps    9;

Qy       3 SCGSPPPILNG---RISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTW 56
           ||  || | ||     |: |     : | |: | |   | | |   : |     ||| |
Db     341 SCAQPPEIPNGEINGAKRVEY--SHGEVVGYDCKPRFLLKGPNKIQC-----VDGMW 380

Qy      57 DKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQA 115
           | | |        || :  | |     || |||| | |   |:| |  ||
Db     381 -TTLPVCVEEERTCGDIPE-LEHGSVKF-SVPFYHHGDSVEFTCAETFTMIGLGSVSC-I 436

Qy     116 NNMWGPTRLPTCVSVFPLE 134
           :  |  | || ||:   ||
Db     437 SGKW--TQLPKCVATDQLE 453


RESULT 23
KFHU13
coagulation factor XIII chain b precursor - human
N;Alternate names: fibrin-stabilizing factor, plasma transglutaminase b chain; protei
C;Species: Homo sapiens (man)
C;Date: 19-Apr-1991 #sequence_revision 26-May-1994 #text_change 22-Jan-1999
C;Accession: A36397; A23830; S09980
R;Bottenus, R.E.; Ichinose, A.; Davie, E.W.
Biochemistry 29, 11195-11209, 1990
A;Title: Nucleotide sequence of the gene for the b subunit of human factor XIII.
A;Reference number: A36397; MUID:91105054
A;Accession: A36397
A;Molecule type: DNA
A;Residues: 1-661 <BOT>
A;Cross-references: GB:J05294
R;Ichinose, A.; McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 25, 4633-4638, 1986
A;Title: Title: Amino acid sequence of the b subunit of human factor XIII, a protein
A;Reference number: A23830; MUID:87026535
A;Accession: A23830
A;Molecule type: mRNA
A;Residues: 2-661 <ICH>
A;Cross-references: GB:M14057; NID:g182838; PIDN:AAA88042.1; PID:g182839
A;Note: parts of this sequence, including the amino end of the mature protein, were d
R;Grundmann, U.; Nerlich, C.; Rein, T.; Zettlmeissl, G.
Nucleic Acids Res. 18, 2817-2818, 1990
A;Title: Complete cDNA sequence encoding the B subunit of human factor XIII.
A;Reference number: S09980; MUID:90251467
A;Accession: S09980
A;Molecule type: mRNA
A;Residues: 1-661 <GRU>
A;Cross-references: EMBL:X51823
C;Comment: This protein is noncatalytic.
C;Genetics:
A;Gene: GDB:F13B
A;Cross-references: GDB:119893; OMIM:134580
A;Map position: 1q31-1q32.1
A;Introns: 2/1; 5/2; 13/1; 19/1; 28/2; 36/2; 92/1; 102/1; 106/1; 56/2; 222/2
C;Complex: heterotetramer in plasma with coagulation factor XIII chain a; the plasma
C;Superfamily: coagulation factor XIII chain b; complement factor H repeat homology
C;Keywords: blood coagulation, duplication, glycoprotein; plasma
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-661/Product: coagulation factor XIII chain b #status predicted <MAT>
F;25-87/Domain: complement factor H repeat homology <FH01>
F;91-146/Domain: complement factor H repeat homology <FH02>
F;153-208/Domain: complement factor H repeat homology <FH03>
F;213-267/Domain: complement factor H repeat homology <FH04>
F;274-327/Domain: complement factor H repeat homology <FH05>
F;336-389/Domain: complement factor H repeat homology <FH06>
F;396-450/Domain: complement factor H repeat homology <FH07>
F;454-515/Domain: complement factor H repeat homology <FH08>
F;524-578/Domain: complement factor H repeat homology <FH09>
F;582-646/Domain: complement factor H repeat homology <FH10>
F;617-619/Region: cell attachment (RGD) motif
F;25-76,59-87,91-135,118-146,153-197,189-208,213-259,241-267,274-315,292-327,340-378
F;162,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             22.0%;  Score 165;  DB 1;  Length 661;
  Best Local Similarity   31.0%;  Pred. No. 5.2e-08;
  Matches   39;  Conservative   19;  Mismatches   58;  Indels   10;  Gaps    5;

Qy       4 CGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPKC 63
           |  ||   :| :|  :||    | ::   |   | : |   : |      || | : |||
Db     274 CPPPPLPINSKIQTHSTTYRHGEIVHIECEINPEIIGSAEIRC---  -EDGKWTEP-PK- 327
```

```
Qy       64 EYENKYSSCPE PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 122
                :  :| | | :  |     |  | :|| ||:|||: : ::(:  : |      | |
Db      328 IESQEKVACEEPPETENGAANLHSKIYYNGDKVIYACKSGYLLHGSNEITCNHGKW--T 384

Qy      123 RLPTCV 128
              | ||
Db      385 LPEECV 390


RESULT 24
I18524
Brain-specific receptor cysteine rich protein homolog sreCM2 - Geodia cydonium
C;Species: Geodia cydonium
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C;Accession: I18524
R;Blumbach, B.; Pancer, Z.; Diehl-Seifert, B.; Steffen, R.; Muenkner, J.; Mueller, I.; M
J. Cell Sci. 111, 2635-2644, 1998
A;Title: The putative sponge aggregation receptor: Isolation and characterisation of a n
A;Reference number: Z18947; MUID:98369060
A;Accession: I18524
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2043 <BLU>
A;Cross-references: EMBL:Y14954; NID:g1364818; PIDN:CAA75175.1; PID:g1364819
C;Genetics:
A;Gene: SRCRM2

  Query Match             21.8%;  Score 164;  DB 2;  Length 2043;
  Best Local Similarity   28.4%;  Pred. No. 2.2e-07;
  Matches   40;  Conservative   25;  Mismatches   56;  Indels   20;  Gaps    6;

Qy        4 CGSPPP       ILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLICITKDKVDGTWD 57
            |   ||          ::  :: |:   : ||: ::|   : | |:| | |::      | |:
Db     1765 CSREPPTKCPTLLISDHVTASSSETTINTVVSPTCDNGYFLKGDKILECLS   PGVWN 1820

Qy       58 KPAPKCEYPNKYSSCPPPIVPGGYKIRGSTPYRHGDSVTFACKTN FSMNGNKSVWCQAN 116
              | |   |   |||  :    :    || :    ||| |  : :::|||| : ||:
Db     1821 GTAITCSLPN   SCPS IKISDHVIANSTIETRINAVVTFTCDDGYTLNGNKTIACQSI 1876

Qy      117 NMWGPT      RLPTCVSVEP 142
              :|  |       :|||  : |
Db     1877 GVWNGTAPTCKEIPICTELIP 1897


RESULT 25
G35070
apolipoprotein H related protein 1Gl - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jul-1998 #sequence_revision 31-Dec-1998 #text_change 31-Mar-2000
C;Accession: G35070; A35071; B35071; H35070; I35070
R;Vik, D.P.; Munoz-Canoves, P.; Koyono, H.; Martin, E.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 4193-4201, 1990
A;Title: Identification and sequence analysis of four complement factor H-related transc
A;Reference number: A35070; MUID:90153969
A;Accession: G35070
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-443 <VIK>
A;Cross-references: GB:M29008; NID:g192559; PIDN:AAA37414.1; PID:g309165; GB:J05259
A;Note: translation of the nucleotide sequence is not complete
C;Superfamily: apolipoprotein H; complement factor H repeat homology
C;Keywords: duplication
F;20-88/Domain: complement factor H repeat homology <FH1>
F;92-145/Domain: complement factor H repeat homology <FH2>
F;154-208/Domain: complement factor H repeat homology <FH3>
F;215-269/Domain: complement factor H repeat homology <FH4>
F;274-334/Domain: complement factor H repeat homology <FH5>

  Query Match             21.6%;  Score 162;  DB 2;  Length 443;
  Best Local Similarity   32.8%;  Pred. No. 4.9e-08;
  Matches   42;  Conservative   16;  Mismatches   60;  Indels   10;  Gaps    5;

Qy        4 CGSPPPILNGPISYYSTPI-AVGTVIPYSCSGTFRLIGEKSLICITKIKVDGTWDKPAPK 62
            || |||| || |:  : | |: |  : : | |   : | | |: :  )       : | | :| |
Db      154 CGPPPPIDNGDITSLSLPVYASESSVEYQCQKYYLEGNKTI C   - RNESWSEP PL 207

Qy       63 CPEYENKYSSCPPPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG  120
            | |       | || |             |   || : :  : : | |:  :| | |
Db      208 CIYPTGKCGSPPPPIDNGDITSISLEYEPESSVEYQCQNYYVLEGEKTIQC NCKWSEP 266

Qy      121 PTRLPTCV 128
            || |  ||
Db      267 PICLSACV 274


RESULT 26
JC5194
membrane cofactor protein precursor - green monkey
N;Alternate names: CD46; MCP; measles virus receptor
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 19-May-2000
C;Accession: JC5194
R;Murakami, Y.; Seya, T.; Kurita, M.; Nagasawa, S.
Biol. Pharm. Bull. 19, 1541-1545, 1996
A;Title: Molecular cloning of a complementary DNA for a membrane cofactor protein (MC
A;Reference number: JC5138; MUID:97149841
A;Contents: Vero cell
A;Accession: JC5194
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-362 <MUR>
A;Cross-references: DDBJ:D78368
C;Comment: This protein is a complement regulator and measles virus receptor. It prot
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C;Keywords: glycoprotein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-362/Product: membrane cofactor protein #status predicted <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;315-336/Domain: transmembrane #status predicted <TMM>
F;347-362/Domain: intracellular #status predicted <INT>
F;83,114,212,214,263,273/Binding site: carbohydrate (Asn) (covalent) #status predicte

  Query Match             21.4%;  Score 160;  DB 2;  Length 362;
  Best Local Similarity   29.1%;  Pred. No. 8.2e-08;
  Matches   39;  Conservative   22;  Mismatches   52;  Indels   16;  Gaps    6;

Qy        2 ISCGSPPPILNGPISYYSTPI-AVGTVIPYSCS  -- GTFRLIGEKSLICITKDKVDGT 55
            | :  || | ||: ::    :      : |||       | ||||  : :      : |
Db      160 ILCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN    NST 215

Qy       56 WDKPAPKCEYPNKYSSCPPPIVPGGYKIRG SPIYRHGDSVTFACKTNFSMNGNKSVWCQ 114
            |   ||:|    |   |  |:|  | :| |    : :  :| | |   : :|||  : |:
Db      216 WSHAAPE---- KVVECPEPVVENGRQISGRGRSFYYEAIVMESVDKGYYLNGSKIVCE 271

Qy      115 ANNMWGPTRLPTCV 128
            :|: | |  :| |:
Db      272 SNSIWDPP VPKCL 284


RESULT 27
JC5138
membrane cofactor protein (MCP, CD46) receptor precursor - green monkey
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C;Accession: JC5138
R;Murakami, Y.; Seya, T.; Kurita, M.; Nagasawa, S.
Biol. Pharm. Bull. 19, 1541-1545, 1996
A;Title: Molecular cloning of a complementary DNA for a membrane cofactor protein (MC
A;Reference number: JC5138; MUID:97149841
A;Contents: Vero cell
A;Accession: JC5138
A;Status: preliminary
```

```
A;Molecule type: mRNA
A;Residues: 1-369 <MUR>
A;Cross-references: DDBJ:D63811; NID:g1777314; PID:g1777315
C;Comment: This protein is a complement regulator and measles virus receptor. It protect
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-369/Product: membrane cofactor protein (MCP, CD46) receptor #status predicted <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;315-336/Domain: transmembrane #status predicted <TMM>
F;347-369/Domain: intracellular #status predicted <INT>

  Query Match             21.4%;  Score 160;  DB 2;  Length 369;
  Best Local Similarity   29.1%;  Pred. No. 8.3e-08;
  Matches   49;  Conservative   22;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS-----GTFRLIGEKSLLCITKDKVDGT 55
           | |  || | ||: ::    :     : |||        | |||| : |      : |
Db     160 ILCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN----NST 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           |   ||:|    |   | |:|  | :| |    : :  :| | |   : :||:  : |:
Db     216 WSHAAPEC----KVVKCREPVVENGKQISGFGKKFYYKATVMFECDKGYYLNGSDKIVCE 271

Qy     115 ANNMWGPTRLPTCV 128
           :|: | |  :| |:
Db     272 SNSTWDPP-VPKCL 284


RESULT 28
154479
membrane cofactor protein precursor, splice form pm5.1 - human
N;Alternate names: lymphocyte surface glycoprotein CD46
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
C;Accession: 154479
R;Purcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.F.
Immunogenetics 44, 335-344, 1991
A;Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator o
A;Reference number: 154479; MUID:91267562
A;Accession: 154479
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-377 <RES>
A;Cross-references: GB:M58050; NID:g180146; PIDN:AAA62833.1; PID:g180137
C;Genetics:
A;Gene: GDB:MCP
A;Cross-references: GDB:120169; OMIM:120920
A;Map position: 1q32-1q32
C;Function:
A;Description: for the factor I mediated cleavage of the complement convertases
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: alternative splicing; glycoprotein; transmembrane protein
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;329-351/Domain: transmembrane #status predicted <TMM>

  Query Match             21.2%;  Score 159;  DB 2;  Length 377;
  Best Local Similarity   27.9%;  Pred. No. 1.1e-07;
  Matches   48;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS-----GTFRLIGEKSLLCITKDKVDGT 55
           | |  || | ||: ::    :     : |||        | |||| :: |      : |
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC----GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | : ||:|    |   | |:|  | :| |    : :  :| | |   | ::|: :: |
Db     216 WSRAAPEC----KVVKCREPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: |    :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 29
S01896
membrane cofactor protein precursor - human
N;Alternate names: lymphocyte surface glycoprotein CD46
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 21-Jul 2000
C;Accession: S01896; A60765; 156188
R;Lublin, D.M.; Liszewski, M.K.; Post, T.W.; Arce, M.A.; Le Beau, M.M.; Rebentisch, M
J. Exp. Med. 168, 181-194, 1988
A;Title: Molecular cloning and chromosomal localization of human membrane cofactor pr
A;Reference number: S01896; MUID:88286080
A;Accession: S01896
A;Molecule type: mRNA
A;Residues: 1-384 <LUB>
A;Cross-references: EMBL:Y00651; NID:g34501; PIDN:CAA68675.1; PID:g34505
A;Note: part of this sequence, including the amino end of the mature protein, was con
R;Purcell, D.F.J.; Deacon, N.J.; Andrew, S.M.; McKenzie, I.F.C.
Immunogenetics 31, 21-28, 1990
A;Title: Human non lineage antigen, CD46 (TRA-2-10): purification and partial sequenci
A;Reference number: A60765; MUID:90129152
A;Accession: A60765
A;Molecule type: protein
A;Residues: 'X',36-39,'X',41-58 <PUR>
R;Cui, W.; Bourcade, D.; Post, T.; Greenlund, A.C.; Atkinson, J.P.; Kumar, V.
J. Immunol. 151, 4137-4146, 1993
A;Title: Characterization of the promoter region of the membrane cofactor protein (CD
A;Reference number: 156188; MUID:94014356
A;Accession: 156188
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-34 <RES>
A;Cross-references: GB:S65879; NID:g425643; PIDN:AAD13968.1; PID:g4261668
C;Genetics:
A;Gene: GDB:MCP
A;Cross-references: GDB:120169; OMIM:120920
A;Map position: 1q32-1q32
C;Function:
A;Description: membrane cofactor protein for the factor I mediated cleavage of the co
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C;Keywords: alternative splicing; glycoprotein; transmembrane protein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-384/Product: membrane cofactor #status experimental <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;329-351/Domain: transmembrane #status predicted <TMM>

  Query Match             21.2%;  Score 159;  DB 2;  Length 384;
  Best Local Similarity   27.9%;  Pred. No. 1.1e-07;
  Matches   48;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS-----GTFRLIGEKSLLCITKDKVDGT 55
           | |  || | ||: ::    :     : |||        | |||| :: |      : |
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC----GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | : ||:|    |   | |:|  | :| |    : :  :| | |   | ::|: :: |
Db     216 WSRAAPEC----KVVKCREPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: |    :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 30
156234
decay accelerating factor - orangutan (fragment)
```

C;Species: Pongo pygmaeus (orangutan)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999
C;Accession: I56234
R;Nickells, M.W.; Alvarez, J.I.; Lublin, D.M.; Atkinson, J.P.
J. Immunol. 152, 676-685, 1994
A;Title: Characterization of DAF 2, a high molecular weight form of decay-accelerating f
A;Reference number: I56234; MUID:94110622
A;Accession: I56234
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-440 <RES>
A;Cross-references: GB:S67775; NID:g459389; PIDN:AAC60609.1; PID:g459390
C;Superfamily: decay accelerating factor; complement factor H repeat homology
F;1-54/Domain: complement factor H repeat homology (fragment) <FH01>
F;57-117/Domain: complement factor H repeat homology <FH02>
F;122-179/Domain: complement factor H repeat homology <FH03>
F;184-242/Domain: complement factor H repeat homology <FH04>

Query Match 21.1%; Score 158.5; DB 2; Length 440;
Best Local Similarity 32.8%; Pred. No. 1.1e-07;
Matches 43; Conservative 20; Mismatches 51; Indels 19; Gaps 8;

Qy 4 SCGSPPPTLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLICITKDKVDGI- -WDKP 59
Db 121 SCPNPGEIRNGQID-VSNGIIFGAIISFSCNTGYKLFGSTSSFCL- -ISGSSVQWSDP 175

Qy 60 APKCEYPNKYSSCPEPIVGGYKIRGSTP YRDGDSVTFACKTNFSMNGNKSVWCQANN 117
Db 176 LPEC REIYCPAPPQIDNGIIQGERDHYGYRQSITYACNKGYTMIGEHSIYCTVND- 231

Qy 118 MW GPTRLPTC 127
Db 232 EGEWSGPP PEC 242

RESULT 31
A26359
decay-accelerating factor, GPI-anchored splice form precursor - human
N;Alternate names: CD55; DAF splice form 2; decay accelerating factor membrane bound for
C;Species: Homo sapiens (man)
C;Date: 05-Oct-1988 #sequence_revision 16-Aug-1996 #text_change 19-Jan-2001
C;Accession: A26359; A27666; A39101; I52594; I52564
R;Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A;Title: Cloning of decay-accelerating factor suggests novel use of splicing to generate
A;Reference number: A26359; MUID:87115845
A;Accession: A26359
A;Molecule type: mRNA
A;Residues: 1-381 <CAR>
A;Cross-references: GB:M30142; NID:g181464; PIDN:AAA52168.1; PID:g181465
R;Medof, M.E.; Lublin, D.M.; Holers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987
A;Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a
A;Reference number: A27666; MUID:87175602
A;Accession: A27666
A;Molecule type: mRNA
A;Residues: 6-79,'T',81-84,'M',86-381 <MED>
A;Cross-references: GB:M15796; NID:g181462; PIDN:AAA52167.1; PID:g181463
R;Moran, P.; Raab, H.; Kohr, W.J.; Caras, I.W.
J. Biol. Chem. 266, 1250-1257, 1991
A;Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag
A;Reference number: A39101; MUID:91093238
A;Accession: A39101
A;Molecule type: protein
A;Residues: 308-352 <MOR>
R;Lublin, D.M.; Mallinson, G.; Poole, J.; Reid, M.E.; Thompson, E.S.; Ferdman, B.K.; Tel
Blood 84, 1276-1282, 1994
A;Title: Molecular basis of reduced or absent expression of decay-accelerating factor in
A;Reference number: I52594; MUID:94326573
A;Accession: I52594
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 191-198,'L',200-209 <LUB>
A;Cross-references: GB:S72858; NID:g639599; PIDN:AAC60633.1; PID:g639600
A;Experimental source: individual KW, Cromer blood group phenotype Dr(a-)
A;Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu
on (see reference I52564), and thus reduced DAF expression
R;Reid, M.E.; Mallinson, G.; Sim, R.B.; Poole, J.; Pausch, V.; Merry, A.H.; Liew, Y.W
Blood 78, 3291-3297, 1991
A;Title: Biochemical studies on red blood cells from a patient with the Inab phenotyp
A;Reference number: I52564; MUID:92075980
A;Accession: I52564
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 190-193,'QLCPVE' <RE2>
A;Cross-references: GB:S70688; NID:g240401; PIDN:AAB20576.1; PID:g240402
A;Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described
C;Comment: Cromer blood group system antigens reside on this protein.
C;Comment: For an alternative splice form, see PIR:A26359
C;Genetics:
A;Gene: GDB:DAF
A;Cross-references: GDB:119088; OMIM:125240
A;Map position: 1q32-1q42
C;Function:
A;Description: protects tissues from damage by regulating complement activation on ce
C;Superfamily: decay-accelerating factor; complement factor H repeat homology
C;Keywords: alternative splicing; blocked carboxyl end; complement inhibitor; glycopr
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-353/Product: decay-accelerating factor 2 #status predicted <MAT>
F;36-94/Domain: complement factor H repeat homology <FH01>
F;98-158/Domain: complement factor H repeat homology <FH02>
F;163-220/Domain: complement factor H repeat homology <FH03>
F;225-283/Domain: complement factor H repeat homology <FH04>
F;354-381/Domain: carboxyl terminal propeptide #status predicted <PRO>
F;95/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature f

Query Match 21.1%; Score 158.5; DB 1; Length 381;
Best Local Similarity 33.4%; Pred. No. 1.2e-07;
Matches 45; Conservative 17; Mismatches 50; Indels 23; Gaps 9;

Qy 4 SCGSPPPTLNGRISYYSTP--IAVGTVIRYSCSGTFRLIGEKSLICITKDKVDGI WD 57
Db 162 SCPNPGEIRNGQI---DVPGGTLFGAIISFSCNTGYKLFGSTSSFCL -- ISGSSVQWS 214

Qy 58 KPAPKCEYPNKYSSCPEPIVGGYKIRGSTP YRDGDSVTFACKTNFSMNGNKSVWCQAN 116
Db 215 EPLPEC REIYCPAPPQIDNGIIQGERDHYGYRQSITYACNKGYTMIGEHSIYCTVN 270

Qy 117 N MW GPTRLPTC 127
Db 271 NDEGEWSGPP PEC 283

RESULT 32
A26359
decay-accelerating factor, splice form 1 precursor - human
N;Alternate names: decay accelerating factor 5; decay accelerating factor secreted fo
C;Species: Homo sapiens (man)
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 02-Jun-2000
C;Accession: A26359; A39702; S16187; S23138; A27258
R;Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A;Title: Cloning of decay-accelerating factor suggests novel use of splicing to gener
A;Reference number: A26359; MUID:87115845
A;Accession: A26359
A;Molecule type: mRNA
A;Residues: 1-440 <CAR>
A;Cross-references: GB:M30142
R;Ewulonu, U.K.; Ravi, L.; Medof, M.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991
A;Title: Characterization of the decay accelerating factor gene promoter region.
A;Reference number: A39702; MUID:91271256
A;Accession: A39702
A;Molecule type: DNA
A;Residues: 1-79,'T',81-104 <EWU>

```
A;Cross references: GB:M64456
A;Note: the authors translated the codon AGT for residue 85 as Met
R;Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.
Biochim. Biophys. Acta 1074, 426-430, 1991
A;Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine.
A;Reference number: S16187; MUID:91291869
A;Accession: S16187
A;Status: preliminary
A;Molecule type: protein
A;Residues: 35-47 <BIO>
R;Nakano, Y.; Sumida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M.
Biochim. Biophys. Acta 1116, 235-240, 1992
A;Title: Complete determination of disulfide bonds localized within the short consensus
A;Reference number: S23138; MUID:92305034
A;Accession: S23138
A;Status: preliminary
A;Molecule type: protein
A;Residues: 35-41;65-68;79-81;93-104;128-134;143-145;155-159;162-168;188-192;203-204;211
R;Sugita, Y.; Nagoro, T.; Matsuda, T.; Sakamoto, T.; Tomita, M.
J. Biochem. 100, 143-150, 1986
A;Title: Improved method for the isolation and preliminary characterization of human DAF
A;Reference number: A27258; MUID:87008461
A;Accession: A27258
A;Molecule type: protein
A;Residues: 35,'X',37,'G',39-51,'P',53-55,'X',57-58,'X',60-63 <SUG>
A;Note: Gly 35 and Leu-38 were also found
C;Comment: For an alternative splice form, see PIR:R26359
C;Genetics:
A;Gene: GDB:DAF
A;Cross references: GDB:119088; OMIM:125240
A;Map position: 1q32-1q32
C;Superfamily: decay-accelerating factor; complement factor H repeat homology
C;Keywords: alternative splicing; glycoprotein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-440/Product: decay-accelerating factor 1 #status predicted <MAT>
F;36-94/Domain: complement factor H repeat homology <FH01>
F;98-158/Domain: complement factor H repeat homology <FH02>
F;163-220/Domain: complement factor H repeat homology <FH03>
F;225-283/Domain: complement factor H repeat homology <FH04>
F;95/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              21.1%;  Score 158.5;  DB 2;  Length 440;
  Best Local Similarity    33.3%;  Pred. No. 1.4e-07;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy        3 SCGSPPPILNGRISYYSTP  IAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT---WD 57
            || :|  | ||:|     |  |  |  | :||:  ::| |  |  |:     : |:   |
Db      162 SCPNPGEIRNGQI---DVPGGILFGATISFSCNTGYKLFGSTSSFCL----ISGSSVQWS 214

Qy       58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSVWCQAN 116
             | |:|    :   || |       |:|    | :  |||:||   |:| |  |::|  |
Db      215 DPLPEC---REIYCPAPPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVN 270

Qy      117 N - MW GPTRLPTC 127
            |    | ||   | |
Db      271 NDEGEWSGPP- PEC 283


RESULT 33
G02913
sperm CD46 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000
C;Accession: G02913
R;Hara, T.
Submitted to GenBank, March 1996
A;Reference number: H01942
A;Accession: G02913
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-349 <HAR>
A;Cross references: GB:D84105; NID:g1256700; PIDN:BAA12224.1; PID:g1256701
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homold
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>

  Query Match              20.6%;  Score 155;  DB 2;  Length 349;
  Best Local Similarity    27.6%;  Pred. No. 2.4e-07;
  Matches   37;  Conservative   25;  Mismatches   56;  Indels   16;  Gaps    6;

Qy        2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCS---GTFRLIGEKSLLCITKDKVDGT 55
            : |  :| | ||: ::    :     : |||        | ||||:: | |      :
Db      160 VLCTPIPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC---GDNSV 215

Qy       56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSVWCQ 114
            | : ||:|    |   |  |:|  | :|     : :   | | |   | ::|: :: |
Db      216 WSRAAPEC----KVVKCRFPVVENGKQISGFGKEFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy      115 ANNMWGPTRLPTCV 128
            :|: | |  :| |:
Db      272 SNSTWDPP-VPKCL 284


RESULT 34
I57998
membrane cofactor protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1995 #sequence_revision 02-Jul-1995 #text_change 19-May-2000
C;Accession: I57998
R;Cervoni, F.; Fenichel, P.; Akhoundi, C.; Hsi, B.L.; Rossi, B.
Mol. Reprod. Dev. 34, 107-113, 1993
A;Title: Characterization of a cDNA clone coding for human testis membrane cofactor p
A;Reference number: I57998; MUID:93119658
A;Accession: I57998
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-369 <RFS>
A;Cross references: GB:S51940; NID:g262937; PIDN:AAB21802.1; PID:g262938
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>

  Query Match              20.6%;  Score 155;  DB 2;  Length 369;
  Best Local Similarity    27.6%;  Pred. No. 2.5e-07;
  Matches   37;  Conservative   25;  Mismatches   56;  Indels   16;  Gaps    6;

Qy        2 ISCGSPPP LNGRISYYSTPI-AVGTVIRYSCS-----GTFRLIGEKSLLCITKDKVDGT 55
            : |  :|  ||: ::    :     : |||        | ||||:: | |      :
Db      160 VLCTPPPK KNGKHTFSEVEVFEYLDAVTYSCDPAPGPLPFSLIGESTIYC----GDNSV 215

Qy       56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
            | : ||:|    |   |  |:|  | :| |    : :  :| | |   | ::|: :: |
Db      216 WSRAAPEC ---KVVKCRFPVVENGKQISG GKEFYYKATVMFECDKGFYLDGSDTIVCD 271

Qy      115 ANNMWGPTRLPTCV 128
            :|: | |  :| |:
Db      272 SNSTWDPP-VPKCL 284


RESULT 35
WMBEIE
secretory complement control protein [validated] precursor - saimiriine herpesvirus 1 (st
N;Alternate names: SCCPH protein
C;Species: saimiriine herpesvirus 1
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 01-Dec-2000
C;Accession: A42534; E36806; S24566
R;Albrecht, J.C.; Fleckenstein, B.
J. Virol. 66, 4937-4940, 1992
A;Title: New member of the multigene family of complement control proteins in herpesv
A;Reference number: A42534; MUID:92260674
A;Accession: A42534
A;Molecule type: DNA
```

```
A;Residues: 1-362 <ALB>
A;Cross-references: EMBL:X60283; NID:g60315; PIDN:CAA42822.1; PID:g60316
C;Genetics:
A;Introns: 288/1
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: alternative splicing; duplication; extracellular protein; glycoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-302/Product: membrane bound complement control protein homolog #status predicted <S
F;23-79/Domain: complement factor H repeat homology <FH1>
F;84-142/Domain: complement factor H repeat homology <FH2>
F;147-205/Domain: complement factor H repeat homology <FH3>
F;210-264/Domain: complement factor H repeat homology <FH4>
F;66-69,46,72,155/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             20.0%;  Score 150;  DB 1;  Length 302;
  Best Local Similarity   29.0%;  Pred. No. 6.1e-07;
  Matches   40;  Conservative   20;  Mismatches   56;  Indels   22;  Gaps    7;

Qy       4 CGSPPPELNGRIS      YYSIPIAVGIVIRYSCSGIFRLIGEKSLIC-IKKVICIW 56
           | :|  :|||| :       |      |:|| |:|:  : |||  :  |: |   :|| |
Db      84 CSTPQDELNGRYTVTGNLYY      GSVITYTCNSGYSLIGSITSACLLKGGRVD-W 135

Qy      57 DKPAPKCEYFNKYSSC-PEPIVPGGYKIRGSTPYRHGDSVTFAC--KTNFSMNGNKSVWC 113
             | |:       | | | :  |        | : |:|||:|  :|  :: |  |  |
Db     136 IPRPPTCD     IKKCKPPQIANGTHTNVKDFYYLDDVTYSCNDGFKLIGSSKLC 191

Qy     114 QANNMWGPTRLPTCVSVF 131
             | |     |  :|
Db     192 SETGSWVPNGPTKCEFIF 209


RESULT 36
WMHBE/F
membrane bound complement control protein homolog precursor - saimiriine herpesvirus 1 (
N;Alternate names: mCCPH protein
C;Species: saimiriine herpesvirus 1
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: B42544; D36806; S24567
R;Albrecht, J.C.; Fleckenstein, B.
J. Virol. 66, 3937-3940, 1992
A;Title: New member of the multigene family of complement control proteins in herpesviru
A;Reference number: A42544; MUID:92260674
A;Accession: B42544
A;Molecule type: DNA
A;Residues: 1-360 <ALB>
A;Cross-references: EMBL:X60283; NID:g60315; PIDN:CAA42823.1; PID:g60317
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-302/Product: membrane bound complement control protein homolog #status predicted <S
F;23-79/Domain: complement factor H repeat homology <FH1>
F;84-142/Domain: complement factor H repeat homology <FH2>
F;147-205/Domain: complement factor H repeat homology <FH3>
F;210-264/Domain: complement factor H repeat homology <FH4>
F;329-345/Domain: transmembrane #status predicted <TMN>
F;66-69,46,72,155,294/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             20.0%;  Score 150;  DB 1;  Length 360;
  Best Local Similarity   29.0%;  Pred. No. 7.4e-07;
  Matches   40;  Conservative   20;  Mismatches   56;  Indels   22;  Gaps    7;

Qy       4 CGSPPPELNGRIS      YYSIPIAVGIVIRYSCSGIFRLIGEKSLIC-IKKVICIW 56
           | :|  :|||| :       |      |:|| |:|:  : |||  :  |: |   :|| |
Db      84 CSTPQDELNGRYTVTGNLYY      GSVITYTCNSGYSLIGSITSACLLKGGRVD-W 135

Qy      57 DKPAPKCEYFNKYSSC-PEPIVPGGYKIRGSTPYRHGDSVTFAC--KTNFSMNGNKSVWC 113
             | |:       | | | :  |        | : |:|||:|  :|  :: |  |  |
Db     136 IPRPPTCD     IKKCKPPQIANGTHTNVKDFYYLDDVTYSCNDGFKLIGSSKLC 191

Qy     114 QANNMWGPTRLPTCVSVF 131
             | |     |  :|
Db     192 SETGSWVPNGPTKCEFIF 209


RESULT 37
A46013
coagulation factor XIII subunit b - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 2-Jun-2000
C;Accession: A46013
R;Nonaka, M.; Matsuda, Y.; Shiroishi, T.; Moriwaki, K.; Nonaka, M.; Matsushima-Sakai, S
Genomics 15, 535-542, 1993
A;Title: Molecular cloning of the b subunit of mouse coagulation factor XIII and assi
A;Reference number: A46013; MUID:93224141
A;Accession: A46013
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-668 <NON>
A;Cross-references: GB:D10071; NID:g303651; PIDN:BAA00963.1; PID:g303652
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:129405, NCBIP:129407)
C;Superfamily: coagulation factor XIII chain b; complement factor H repeat homology
F;25-87/Domain: complement factor H repeat homology <FH01>
F;91-146/Domain: complement factor H repeat homology <FH02>
F;153-208/Domain: complement factor H repeat homology <FH03>
F;213-267/Domain: complement factor H repeat homology <FH04>
F;274-327/Domain: complement factor H repeat homology <FH05>
F;336-389/Domain: complement factor H repeat homology <FH06>
F;396-450/Domain: complement factor H repeat homology <FH07>
F;454-515/Domain: complement factor H repeat homology <FH08>
F;524-578/Domain: complement factor H repeat homology <FH09>
F;582-646/Domain: complement factor H repeat homology <FH10>

  Query Match             19.8%;  Score 149;  DB 2;  Length 668;
  Best Local Similarity   31.7%;  Pred. No. 1.8e-06;
  Matches   41;  Conservative   16;  Mismatches   60;  Indels   10;  Gaps    5;

Qy       4 CGSPPPELNGRISYYSIPIAVGIVIRYSCSGIFRLIGEKSLIC-IKKVICIWDKPAPKC 63
           |  ||  || :|  :||    |  :   |   | : | :|||       :| | :| |||
Db     274 CPPPPVPLNSKIQPHSTTYRHCERVHIECELNFVIQCSEELIC-     ENGKWIEP-PKC 327

Qy      64 EYFNKYSSCPE-PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 122
           :  :| : | |  |      | |  || ||: |   :|: |: :: |     | |
Db     328 IEEKEKVACPQPPSVENGVAHPHSFIYYSCQKVTYRCGGGYSLPSSSTITCN PGRW T 384

Qy     123 RLPTCV 128
             | :|
Db     385 LPPECV 390


RESULT 38
A42755
P-selectin precursor - mouse
N;Alternate names: CD62; granule membrane protein 140; PADGEM
C;Species: Mus musculus (house mouse)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 19-May-2000
C;Accession: A42755; A44899
R;Weller, A.; Isenmann, S.; Vestweber, D.
J. Biol. Chem. 267, 15176-15183, 1992
A;Title: Cloning of the mouse endothelial selectins. Expression of both E- and P-sele
A;Reference number: A42755; MUID:92340571
A;Accession: A42755
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-768 <WEL>
A;Cross-references: GB:M87861; NID:g200552; PIDN:AAA40008.1; PID:g200553
A;Experimental source: endothelial cells
A;Note: sequence extracted from NCBI backbone (NCBIP:109467)
R;Sanders, W.E.; Wilson, R.W.; Ballantyne, C.M.; Beaudet, A.L.
Blood 80, 795-800, 1992
A;Title: Molecular cloning and analysis of in vivo expression of murine P-selectin.
A;Reference number: A44899; MUID:92345617
A;Accession: A44899
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-723,'E',725-768 <SAN>
```

```
A;Cross references: GB:M72432; NID:g193565; PIDN:AAA37712.1; PID:g193566
A;Note: sequence extracted from NCBI backbone (NCBIP:109900)
C;Superfamily: unassigned EGF-related proteins; complement factor H repeat homology; ECH
C;Keywords: cell adhesion; glycoprotein; phosphohistidine; phosphoprotein; transmembrane
F;1-41/Domain: signal sequence #status predicted <SIG>
F;42-768/Product: P-selectin #status predicted <MAT>
F;163-194/Domain: EGF homology <EGF>
F;209-257/Domain: complement factor H repeat homology <FH01>
F;262-319/Domain: complement factor H repeat homology <FH02>
F;324-381/Domain: complement factor H repeat homology <FH03>
F;386-443/Domain: complement factor H repeat homology <FH04>
F;448-505/Domain: complement factor H repeat homology <FH05>
F;510-567/Domain: complement factor H repeat homology <FH06>
F;580-637/Domain: complement factor H repeat homology <FH07>
F;642-699/Domain: complement factor H repeat homology <FH08>
F;710-733/Domain: transmembrane #status predicted <TMN>
F;734-768/Domain: intracellular #status predicted <INT>
F;45,54,107,212,347,398,456,467,603,654,661,679/Binding site: carbohydrate (Asn) (covale

  Query Match             19.7%;  Score 148;  DB 2;  Length 768;
  Best Local Similarity   28.6%;  Pred. No. 2.6e-06;
  Matches   86;  Conservative   21;  Mismatches   49;  Indels   20;  Gaps    6;

Qy      16 SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPKCEYFNKYSS---- 71
           |:      ||::  :||:  | ||:| ::: |     | | |  | | |:      :
Db     524 SHVHGFFGVGSICHFSCNEDFELIGSENVFC----TVSGRWSAPPPTCKGITSLPAPAVR 579

Qy      72 CPEPIVPGGYKIRGSTPYRH---GDSVI-FACKTNFSMNGNKSVWCQANNMWGPTRL 124
           ||    ||    :||:  :|     | : |  | ||| |:: |  |: |:|:  |
Db     580 CPAITTPG---QGTMSCQHHEGSFGFNTTCYFGCKTGFTLRGANSIRCRASGQWTAV-T 634

Qy     125 PTCVSV 130
           | | :|
Db     635 PMCRAV 640


RESULT 39
I56100
complement factor H related protein 1 precursor - human
N;Alternate names: factor H homolog; FHR-1
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I56100; I72654; A40455; S14604
R;Estaller, C.; Koistinen, V.; Schwaeble, W.; Dierich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A;Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel

A;Reference number: I56100; MUID:91201892
A;Accession: I56100
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-330 <EST1>
A;Cross-references: GB:M65292; NID:g183762; PIDN:AAA35946.1; PID:g183763
A;Accession: I72653
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-156,'H',158,'L',160-174,'E',176-330 <EST2>
A;Cross-references: GB:M65293; NID:g183764; PIDN:AAA35947.1; PID:g183765
A;Skerka, C.; Horstmann, R.D.; Zipfel, P.F.
J. Biol. Chem. 266, 12015-12020, 1991
A;Title: Molecular cloning of a human serum protein structurally related to complement F
A;Reference number: A40455; MUID:91268081
A;Accession: A40455
A;Molecule type: mRNA
A;Residues: 4-70,'N',72-330 <SKE>
A;Cross-references: EMBL:X56209; NID:g31990; PIDN:CAA39666.1; PID:g388519
C;Genetics:
A;Gene: GDB:HFL1; H36-1; FHR1
A;Cross-references: GDB:128809; OMIM:134371
C;Superfamily: apolipoprotein H; complement factor H repeat homology
C;Keywords: duplication; extracellular protein; glycoprotein; tandem repeat
F;1-18/Domain: signal sequence #status predicted <SIG>
F;24-84/Domain: complement factor H repeat homology <FH1>
F;87-140/Domain: complement factor H repeat homology <FH2>
F;147-201/Domain: complement factor H repeat homology <FH3>
F;208-262/Domain: complement factor H repeat homology <FH4>
F;266-327/Domain: complement factor H repeat homology <FH5>

  Query Match             19.6%;  Score 147;  DB 2;  Length 430;
  Best Local Similarity   27.9%;  Pred. No. 1.4e-05;
  Matches   36;  Conservative   22;  Mismatches   61;  Indels   10;  Gaps    5;

Qy       3 SCGSPPPIINGRI-SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           || :|| : |  | |   :    |  :|| |   : : |:  :::|     :| | :| |
Db     146 SCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCL-- - NGNWTEP P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVIFACKTNFSMNGNKSVWCQANNMWG- 120
           :|:       | || |       : |    |  : |:  : : ||| : |:  |  |
Db     200 QCKDSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NGQWSE 258

Qy     121 -PTRLPTCV 128
            |  |  ||
Db     259 PPKCLHPCV 267


RESULT 40
I53821
P-selectin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 19-May-2000
C;Accession: I53821
R;Auchampach, J.A.; Oliver, M.G.; Anderson, D.C.; Manning, A.M.
Gene 145, 251-255, 1994
A;Title: Cloning, sequence comparison and in vivo expression of the gene encoding rat
A;Reference number: I53821; MUID:94333817
A;Accession: I53821
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-768 <RES>
A;Cross-references: GB:L23088; NID:g349552; PIDN:AAA60825.1; PID:g349553
C;Superfamily: unassigned EGF-related proteins; C-type lectin homology; complement fa
F;32-158/Domain: C-type lectin homology <LCH>
F;163-194/Domain: EGF homology <EGF>
F;262-319/Domain: complement factor H repeat homology <FHR>
F;510-567/Domain: complement factor H repeat homology <FH06>
F;580-637/Domain: complement factor H repeat homology <FH07>
F;642-699/Domain: complement factor H repeat homology <FH08>

  Query Match             19.5%;  Score 146.5;  DB 2;  Length 768;
  Best Local Similarity   28.7%;  Pred. No. 3.6e-06;
  Matches   43;  Conservative   20;  Mismatches   54;  Indels   33;  Gaps    9;

Qy       5 GSPP---  ---PIL----NGRI--SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDK 51
           ||||       | :     | :   |:     :||:   :||   | |:| ::: |
Db     500 GSPPMCEAIKCPEIFAPEQGSLDCSHVHGEFSVSSTCHFSCNEEFELLGSRNVFC----1 555

Qy      52 VDGTWDKPAPKCEYFNKYS----SCPEPIVPGGYKIRGSTPYRH-   --GDSVI--FACK 100
           | | |    | |:            ||    ||    :|:   ||      | :  |  ||
Db     556 VSGRWSAPPPTCKGVTSLPVPSVFCPALITPG - -QGIMSC RHHLESFGFNITCYFG K 611

Qy     101 TNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           | |::  |  :| |:  |     | | :|
Db     612 TGFTLRGANSIRCGASGQWTAV-TPVCPAV 640


RESULT 41
A38738
coagulation factor C precursor - horseshoe crab (Tachypleus tridentatus)
N;Alternate names: coagulation-complement factor C; Limulus factor C
N;Contains: coagulation factor C heavy chain; coagulation factor C light chain peptid
C;Species: Tachypleus tridentatus
C;Date: 04-Oct-1991 #sequence_revision 04-Oct-1991 #text_change 08-Dec-2000
C;Accession: A38738; B38738; S00105
R;Muta, T.; Miyata, T.; Misumi, Y.; Tokunaga, F.; Nakamura, T.; Toh, Y.; Ikehara, Y.
J. Biol. Chem. 266, 6554-6561, 1991
A;Title: Limulus factor C. An endotoxin-sensitive serine protease zymogen with a mosa
```

A;Reference number: A38748; MUID:91177916
A;Accession: A38748
A;Molecule type: mRNA
A;Residues: 1-1019 <MUI>
A;Cross-references: GB:D90271; NID:g217396; PID:d1015020; PID:g217397
A;Accession: B38748
A;Molecule type: mRNA
A;Residues: 1-466,616,'DN',619-620,'A',622 <MU2>
A;Cross-references: GB:D90272
R;Iokubaqa, K.; Miyata, T.; Nakamura, I.; Morita, T.; Kuma, K.I.; Miyata, T.; Iwanaga, S
Eur. J. Biochem. 167, 405-416, 1987
A;Title: Lipopolysaccharide sensitive serine protease zymogen (factor C) of horseshoe cr
is a novel type of serine protease.
A;Reference number: S00105; MUID:88004461
A;Accession: S00105
A;Molecule type: protein
A;Residues: 'S',7-89,'S',41,'I',43-54,'E';691-782;950-977 <IOK>
C;Superfamily: coagulation factor C; C type lectin homology; complement factor H repeat
C;Keywords: alternative splicing; glycoprotein; hemolymph coagulation; hydrolase; serine
F;1-15/Domain: signal sequence #status predicted <SIG>
F;25-690/Product: coagulation factor C heavy chain #status experimental <HCH>
F;130-190/Domain: complement factor H repeat homology <FH01>
F;199-254/Domain: complement factor H repeat homology <FH02>
F;260-321/Domain: complement factor H repeat homology <FH03>
F;439-504/Domain: C type lectin homology <LCH>
F;526-584/Domain: complement factor H repeat homology <FH04>
F;689-744/Domain: complement factor H repeat homology #status atypical <FH05>
F;695-762/Product: coagulation factor C light chain peptide A #status experimental <PPA>
F;764-1019/Product: coagulation factor C light chain peptide B #status experimental <PPB
F;764-1015/Domain: trypsin homology <TRY>
F;524,584,624,912/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;767/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;809,865,960/Active site: His, Asp, Ser #status predicted

  Query Match 18.0%; Score 142; DB 2; Length 1019;
  Best Local Similarity 24.3%; Prod. No. 1.4e-05;
  Matches 98; Conservative 41; Mismatches 50; Indels 36; Gaps 8;

Qy 2 ISC-GSPPPILNGR-ISYYSTPIAVGTVIRYSCSGTFRLIGEKSL 44
Db 186 ISCTLPNGQWSSFPPKCIREC-AKVSSPEHGKVNAPSGNMIEGATLRFSCDSPYYLEGPTL 239

Qy 45 LTIKDKVDGIWDPEIAESCIYENKYSSCP EPIVPGGYKIR GSIPYRHGDSV 95
Db 240 IC - QGNGQWSGQIPQCK KLVECPDLDPVNHAERQVKIGVEQKYGQFP QGTEV 290

Qy 96 IFACKTNFSMNGNKSVWSQANNMWCPIVPTCVSV 130
Db 291 IYICSGNYFIMGENTIKCNPDGSWSGSQ PSCVKV 324

RESULT 4
A34877
C4b binding protein beta chain precursor - human
N Alternate names: protein S binding chain, C4b
C;Species: Homo sapiens (man)
C;Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 21-Jul-2000
C;Accession: A47107; A34877; A31117
R;Hillarp, A.; Fardo Manuel, F.; Ruiz, R.R.; Rodriguez de Cordoba, S.; Dahlbaeck, B.
J. Biol. Chem. 268, 15017-15023, 1993
A;Title: The human C4b binding protein beta chain gene.
A;Reference number: A47107; MUID:93315479
A;Accession: A47107
A;Molecule type: DNA
A;Residues: 1-252 <HIL>
A;Cross-references: GB:L11244; NID:g179682; PID:g387660; GB:L11245; NID:g179683; PID:g17
R;Hillarp, A.; Dahlbaeck, B.
Proc. Natl. Acad. Sci. U.S.A. 87, 1183-1187, 1990
A;Title: Cloning of cDNA coding for the beta chain of human complement component C4b-bin
A;Reference number: A34877; MUID:90138962
A;Accession: A34877
A;Molecule type: mRNA
A;Residues: 1-252 <HIL>
A;Cross-references: GB:M29964; NID:g179678; PIDN:AAB59521.1; PID:g179679
R;Hillarp, A.; Dahlbaeck, B.
J. Biol. Chem. 263, 12759-12764, 1988
A;Title: Novel subunit in C4b binding protein required for protein S binding.
A;Reference number: A31117; MUID:88315081
A;Accession: A31117
A;Molecule type: protein
A;Residues: 18-20,'PQ',23,26-27,'L',29,'P',31,'VD',34-36,'F' <HIS>
C;Genetics:
A;Gene: GDB:C4BPB
A;Cross-references: GDB:125208; OMIM:120831
A;Map position: 1q32-1q32
C;Superfamily: complement factor H repeat homology
C;Keywords: extracellular protein; plasma
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-252/Product: C4b-binding protein beta chain #status experimental <MAT>
F;23-76/Domain: complement factor H repeat homology <FH01>
F;81-134/Domain: complement factor H repeat homology <FH02>
F;139-191/Domain: complement factor H repeat homology <FH03>

  Query Match 18.8%; Score 141; DB 2; Length 252;
  Best Local Similarity 26.0%; Prod. No. 3.7e-06;
  Matches 38; Conservative 21; Mismatches 59; Indels 14; Gaps 6;

Qy 4 CGSPPPILNGR-ISYYSTPIAVGTVIRYSCSGTFRLIGEKSLICIKDKVDGIWDPEPAPK 62
Db 23 CPELPPVQN-- SIEVAKFVFGQTITGTYVCTKGYRLVGKKTLFCI NASKEWDNTTIE 75

Qy 63 CSEFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWSQANNMWCPT 122
Db 76 C KLGEVFDFVLVNG EFSSSGFVNVSDKLTFMCNDHYTLKGSNKSQCLEDHTWAPP 130

Qy 123 KLPPCVS 129
Db 131 FPTCKS 136

RESULT 43
A53274
complement factor B subunit Ba - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 02-May-1994 #sequence_revision 18-Nov-1994 #text_change 19-May-2000
C;Accession: A53274
R;Peelman, L.J.; Van de Weghe, A.R.; Coppieters, W.R.; Van Zeveren, A.J.; Bouquet, Y.
Immunogenetics 34, 192-195, 1991
A;Title: Cloning and sequencing of the porcine complement factor B.
A;Reference number: A53274; MUID:91372866
A;Accession: A53274
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <PEE>
A;Note: sequence extracted from NCBI backbone (NCBIN:56029, NCBIP:56042)
C;Superfamily: complement C2; complement factor B repeat homology; trypsin homology;
F;4-59/Domain: complement factor B repeat homology <FH1>
F;66-119/Domain: complement factor B repeat homology <FH2>

  Query Match 18.6%; Score 140; DB 2; Length 152;
  Best Local Similarity 30.6%; Prod. No. 2.7e-06;
  Matches 41; Conservative 15; Mismatches 62; Indels 16; Gaps 6;

Qy 2 ISCGSPPPILNG--- RISYYSTPIAVGTVIRYSCSGTFELIGEKSLICIKDKVDGIWD 57
Db 2 IRCPRPHDPENGEYWPRAPYYN ---LSDEISFHCYEGYTLRGSANRTC QVIGRWD 53

Qy 58 KPAPKCSEFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWSQANN 117
Db 54 GQTAICDDGAGY--CPNPGIPIGTRKVG TQYRLEDSVTYYCTRGLTLRGSQRKICQEGG 110

Qy 118 MWCPIRLPTCVSVF 131
Db 111 SWSGIE PSCQDSF 123

```
RESULT 44
S57960
C4BP protein beta chain precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 05-Nov-1999
C;Accession: S57960
R;Hillarp, A.; Thern, A.; Dahlback, B.
submitted to the EMBL Data Library, July 1995
A;Description: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: stru
A;Reference number: S57954
A;Accession: S57960
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-258 <HIL>
A;Cross-references: EMBL:Z50052; NID:g899381; PIDN:CAA90392.1; PID:g899382
C;Superfamily: complement factor H repeat homology
F;20-74/Domain: complement factor H repeat homology <FH01>
F;78-141/Domain: complement factor H repeat homology <FH02>
F;146-188/Domain: complement factor H repeat homology <FH?>

  Query Match              18.6%;  Score 140;  DB 2;  Length 258;
  Best Local Similarity    27.1%;  Pred. No. 4.7e-06;
  Matches   46;  Conservative   18;  Mismatches   55;  Indels   24;  Gaps    6;

Qy       8 SCGSPPPILNGRISYYSTPIAVGTVIR----YSCSGTFRLIGEKSLLCITKDKVDGTW 56
           ||  |||: |        : ||          | |   : |:|:|||:     :    |
Db      19 SCSEPPPVNNS------VEVGKEIEEQTLGIYILCIKGYHLVGKKSLVFDPSKE---W 66

Qy      57 DKPAPKCEYFNKYSSCPPPIVDGGYKTRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQAN 116
           :   |:|        ||:|::  | ||  | |      : | |   : : |:    |  :
Db      67 NSTLPFC--LIGHCPDPVIENG KINSSGPVNISGKTMFFCNDGYILKGSNWSQCLED 121

Qy     117 NMWCPTRLPTCVS 129
           : | |  ||: | |
Db     122 HTWAPP-LPICRS 133


RESULT 45
C2HU
complement C2 precursor [validated] - human
N;Contains: classical complement pathway C3/C5 convertase (EC 3.4.21.43) C2a subunit; cl
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Mar-2001
C;Accession: A25971; JS0281; A25290; A05289; A37539, A37540, A26506, B26506, I56179, I54
R;Wu, L.; Morley, B.J.; Campbell, R.D.
Cell 48, 841-842, 1987
A;Title: Cell-specific expression of the human complement protein factor B gene: evidenc
A;Reference number: A25971; MUID:87102880
A;Accession: A25971
A;Molecule type: DNA
A;Residues: 694-752 <WUL>
A;Cross-references: GB:M15882; NID:g182699; PIDN:AAA59624.1; PID:g467309
R;Horiuchi, T.; Macon, K.J.; Kidd, V.J.; Volanakis, J.E.
J. Immunol. 142, 2105-2111, 1989
A;Title: cDNA cloning and expression of human complement component C2.
A;Reference number: JS0281; MUID:89156483
A;Accession: JS0281
A;Molecule type: mRNA
A;Residues: 1-542,'F',534-752 <HOR>
A;Cross-references: GB:M26091
R;Bentley, D.R.
Biochem. J. 239, 339-345, 1986
A;Title: Primary structure of human complement component C2.
A;Reference number: A25290; MUID:87127920
A;Accession: A25290
A;Molecule type: mRNA
A;Residues: 1-752 <BEN>
A;Cross-references: GB:X04481; NID:g34627; PIDN:CAA28169.1; PID:g34628
R;Bentley, D.R.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1212-1215, 1984
A;Title: Isolation of cDNA clones for human complement component C2.
A;Reference number: A05289; MUID:84144868
A;Accession: A05289
A;Molecule type: mRNA
A;Residues: 588-752 <BE2>
A;Cross-references: GB:K01236
R;Cagnon, J.
Phil. Trans. R. Soc. Lond. B Biol. Sci. 306, 301-309, 1984
A;Title: Structure and activation of complement components C2 and factor B.
A;Reference number: A37539; MUID:85038851
A;Accession: A37539
A;Molecule type: protein
A;Residues: 137-149,'AG',150-171;454-466;574-717 <GAG>
A;Note: two glycosylation sites were determined
R;Parkes, C.; Cagnon, J.; Kerr, M.A.
Biochem. J. 213, 201-209, 1983
A;Title: The reaction of iodine and thiol-blocking reagents with human complement com
thiol group.
A;Reference number: A37540; MUID:83308588
A;Accession: A37540
A;Molecule type: protein
A;Residues: 244-267,'G',269 <PAR>
R;Kerr, M.A.; Cagnon, J.
Biochem. J. 205, 59-67, 1982
A;Title: The purification and properties of the second component of guinea pig comple
A;Reference number: A26506; MUID:83022289
A;Accession: A26506
A;Molecule type: protein
A;Residues: 244-248,'SX',251,'XK',254-256 <KER>
A;Accession: B26506
A;Molecule type: protein
A;Residues: 21-23,'X',25-28,'XLX',32-33,'S',35-39,'X',40,'X',42-44,'X',46 <KE2>
R;Ishii, Y.; Zhu, Z.B.; Macon, K.J.; Volanakis, J.E.
J. Immunol. 151, 170-174, 1993
A;Title: Structure of the human C2 gene.
A;Reference number: I56179; MUID:93315833
A;Accession: I56179
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-512,'F',534-752 <RES>
A;Cross-references: GB:L09708; NID:g179654; PIDN:AAB97607.1; PID:g298124
R;Bentley, D.R.; Campbell, R.D.; Cross, S.
Immunogenetics 22, 377-390, 1985
A;Title: DNA polymorphism of the C2 locus
A;Reference number: I54419; MUID:86032058
A;Accession: I54419
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 21-46 <RE2>
A;Cross-references: GB:M15549; NID:g187764; PIDN:AAA59649.1; PID:g187765
C;Genetics:
A;Gene: GDB:C2
A;Cross-references: GDB:119731; OMIM:217000
A;Map position: 6p21.3-6p21.3
A;Introns: 16/1, 80/1, 148/1, 206/1, 249/3, 283/3, 335/1, 377/1, 435/3, 475/3, 1805/3;
C;Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C
C;Function:
A;Description: cleaves complement C3 and complement C5 alpha chains
A;Pathway: complement classical pathway
C;Superfamily: complement C2; complement factor H repeat homology; trypsin homology;
C;Keywords: complement classical pathway; duplication; glycoprotein; hydrolase; plasm
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-243/Product: complement C2b subunit #status experimental <C2B>
F;24-84/Domain: complement factor H repeat homology <FH1>
F;89-144/Domain: complement factor H repeat homology <FH2>
F;151-204/Domain: complement factor H repeat homology <FH3>
F;244-752/Product: complement C2a subunit #status experimental <C2A>
F;252-442/Domain: von Willebrand factor type A repeat homology <VFA>
F;471-739/Domain: trypsin homology #status atypical <TRY>
F;24-64,51-84,89-141,117-144,151-191,177-204,464-541,492-538,584-630,668-665,675-717/
F;29,112,290,333,467,471/Binding site: carbohydrate (Asn) (covalent) #status predicte
F;243-244/Cleavage site: Arg-Lys (complement subcomponent C2a) #status experimental
F;507,561,679/Active site: His, Asp, Ser #status predicted
F;621,651/Binding site: carbohydrate (Asn) (covalent) #status experimental

  Query Match              18.6%;  Score 140;  DB 1;  Length 752;
```

```
  Best Local Similarity   28.5%;  Pred. No. 1.5e-05;
  Matches   47;  Conservative   16;  Mismatches   61;  Indels   16;  Gaps    6;

Qy        7 ISCGSPPPILNG    RTSYYSTPTAVCIVTRYSCSGIFRLIGEKSLICTIKDKVDGIWD 57
            : | :     ||    |   |      ||  : : |   | | |     |    : :| ||
Db       97 VRCPAPVSPENGIYIPRLGSY    PVGGNVSFECEDGFTLRGSPVRQC -  RPNGMWD 138

Qy       58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
                :  |    || , :  :  :|    : ||| | : | :|  : |:    || |
Db      139 GETAVCD  NGAGHCPNPCISLC AVRTGFRFGHGDKVRYRCSSNLVLTGSSERECQGNG 195

Qy      118 MWGPTELPIC 127
            :|  |  | |
Db      196 VWSGTE PIC 204


Search Completed: November  6, 2002, 16:07:30
Job Time   19.4461 secs
```

GenCore version 5.1.3

OM protein protein search, using sw model

Run on: November 6, 2002, 16:04:58 ; Search time 8.4684 Seconds
(without alignments)
612.680 Million cell updates/sec

Title: US-09-834-309-4
Perfect score: 751
Sequence: 1 GISCGSPPPIINGRISYYST ANNMWGPTRLPTCVSVFPLE 134

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 48719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	751	100.0	1033	1	CR2_HUMAN	P20023 homo sapien
2	466.5	62.1	1025	1	CR2_MOUSE	P19070 mus musculu
3	219	29.2	2039	1	CR1_HUMAN	P17927 homo sapien
4	215	28.6	263	1	VCP_VACCV	P10998 vaccinia vi
5	199	26.5	597	1	C4BP_HUMAN	P04003 homo sapien
6	197.5	26.3	610	1	C4BP_BOVIN	Q28065 bos taurus
7	182	24.2	558	1	C4BP_RAT	Q63514 rattus norv
8	180.5	24.0	1231	1	CFAH_HUMAN	P08603 homo sapien
9	178.5	23.8	469	1	C4BP_MOUSE	P08607 mus musculu
10	178.5	23.8	1234	1	CFAH_MOUSE	P06909 mus musculu
11	167.5	22.3	390	1	DAF1_MOUSE	Q61475 mus musculu
12	165	22.0	661	1	FI3B_HUMAN	P05160 homo sapien
13	161	21.4	407	1	DAF2_MOUSE	Q61476 mus musculu
14	159	21.2	377	1	MCP_HUMAN	P15529 homo sapien
15	158.5	21.1	440	1	DAF_PONPY	P49457 pongo pygma
16	158.5	21.1	481	1	DAF_HUMAN	P08174 homo sapien
17	150	20.0	360	1	CCPH_HSVSA	Q01016 herpesvirus
18	149	19.8	668	1	FI3B_MOUSE	Q07968 mus musculu
19	148	19.7	768	1	LEM3_MOUSE	Q01102 mus musculu
20	147	19.6	330	1	FHR1_HUMAN	Q03591 homo sapien
21	146.5	19.5	768	1	LEM3_RAT	P98106 rattus norv
22	144	19.2	1019	1	LEC_CAERE	Q26422 caenorhabd
23	142	18.9	1019	1	LEC_TACTR	P28175 tachypleus
24	141.5	18.8	507	1	DAF_CAVPO	Q60401 cavia porce
25	141	18.8	252	1	C4BB_HUMAN	P20851 homo sapien
26	140	18.6	151	1	CFAB_PIG	Q03710 sus scrofa
27	140	18.6	258	1	C4BB_RAT	Q63515 rattus norv
28	140	18.6	752	1	CO2_HUMAN	P06681 homo sapien
29	137	18.2	345	1	APOH_MOUSE	Q01339 mus musculu
30	136.5	18.2	958	1	BIG_DROME	Q09101 drosophila
31	134.5	17.9	270	1	FHR2_HUMAN	P36980 homo sapien
32	134	17.8	297	1	APOH_RAT	P26644 rattus norv
33	133	17.7	764	1	CFAB_HUMAN	P00751 homo sapien
34	132	17.6	345	1	APOH_CANFA	P33704 canis famil
35	131	17.4	202	1	APAR_PIG	Q03472 sus scrofa
36	131	17.4	685	1	CFAH_BOVIN	Q28085 bos taurus
37	131	17.4	769	1	LEM3_SHEEP	P98109 ovis aries
38	129.5	17.2	549	1	LEM2_RAT	P98105 rattus norv
39	127	16.9	345	1	APOH_BOVIN	P17690 bos taurus
40	127	16.9	345	1	APOH_HUMAN	P02749 homo sapien
41	127	16.9	646	1	LEM3_BOVIN	P42201 bos taurus
42	123.5	16.4	612	1	LEM2_MOUSE	Q06890 mus musculu
43	123	16.4	934	1	CO6_HUMAN	P13671 homo sapien
44	122	16.2	611	1	LEM2_CANFA	P33730 canis famil
45	122	16.2	761	1	CFAB_MOUSE	P04186 mus musculu
46	120.5	16.0	760	1	CO2_MOUSE	P21180 mus musculu
47	120	16.0	610	1	LEM2_HUMAN	P15502 homo sapien
48	119	15.8	331	1	FHR4_HUMAN	Q92496 homo sapien
49	118	15.7	330	1	FHR3_HUMAN	Q02985 homo sapien
50	118	15.7	484	1	LEM2_PIG	P98110 sus scrofa
51	117.5	15.6	551	1	LEM2_RABIT	P27113 oryctolagus
52	114	15.2	830	1	LEM3_HUMAN	P16109 homo sapien
53	110.5	14.7	843	1	CO7_HUMAN	P10643 homo sapien
54	110	14.6	485	1	LEM2_BOVIN	P98107 bos taurus
55	102.5	13.6	317	1	VB05_VACCC	P24083 vaccinia vi
56	102	13.6	464	1	SRPX_RAT	Q63769 rattus norv
57	101.5	13.5	686	1	MAS2_HUMAN	O00187 homo sapien
58	99.5	13.2	317	1	VB05_VACCO	P24084 vaccinia vi
59	94.5	12.6	317	1	VB05_VACCV	Q01227 vaccinia vi
60	94	12.5	464	1	SRPX_HUMAN	P78539 homo sapien
61	93	12.4	347	1	HPT_RAT	P06866 rattus norv
62	90.5	12.1	317	1	VB05_VACCI	P21115 vaccinia vi
63	86	11.5	699	1	CRAR_HUMAN	P48740 n complemen
64	85.5	11.4	926	1	PERT_PIG	P09933 sus scrofa
65	85	11.3	198	1	C4BB_BOVIN	Q28066 bos taurus
66	85	11.3	347	1	HPT_MUSSA	Q62558 mus saxicol
67	84	11.2	370	1	LEM1_BOVIN	P98131 bos taurus
68	83.5	11.1	406	1	HPT2_HUMAN	P00738 homo sapien
69	83	11.1	250	1	CFBL_CHICK	P81475 gallus gall
70	83	11.1	933	1	PERT_HUMAN	P07202 homo sapien
71	82.5	11.0	810	1	PLMN_MACMU	P12545 macaca mula
72	81	10.8	372	1	LEM1_MACMU	Q95198 macaca mula
73	81	10.8	372	1	LEM1_PAPHA	Q28768 papio hamad
74	80	10.7	372	1	LEM1_HUMAN	P24151 homo sapien
75	80	10.7	372	1	LEM1_PANTR	Q95237 pan troglod
76	79	10.5	704	1	CRAR_MOUSE	P98064 mus musculu
77	78.5	10.5	1627	1	PAPA_HUMAN	Q13219 homo sapien
78	78.5	10.5	3562	1	PGCV_CHICK	Q90953 gallus gall
79	77	10.3	372	1	LEM1_PONPY	Q95235 pongo pygma
80	74	9.9	347	1	HPT_ATECE	P50417 ateles geof
81	73.5	9.8	470	1	NRAM_I72A1	P11485 influenza a
82	73.5	9.8	662	1	MM02_MOUSE	P33436 mus musculu
83	73.5	9.8	662	1	MM02_RAT	P33436 rattus norv
84	73.5	9.8	684	1	XYNA_CALSR	P40944 caldicellul
85	73.5	9.8	812	1	PLMN_BOVIN	P06868 bos taurus
86	72.5	9.7	378	1	CXA1_XENLA	P16863 xenopus lae
87	72.5	9.7	470	1	NRAM_I41N9	P41349 influenza a
88	72.5	9.7	912	1	PGCB_BOVIN	Q28062 bos taurus
89	71.5	9.5	470	1	NRAM_I49AE	P31348 influenza a
90	71.5	9.5	628	1	LU_HUMAN	P50895 homo sapien
91	71.5	9.5	688	1	C1S_HUMAN	P09871 homo sapien
92	71.5	9.5	705	1	C1R_HUMAN	P00736 homo sapien
93	71.5	9.5	883	1	PGCB_MOUSE	Q61361 mus musculu
94	71.5	9.5	883	1	PGCB_RAT	P55068 rattus norv
95	71	9.5	812	1	PLMN_MOUSE	P20918 mus musculu
96	71	9.5	1700	1	BAR3_CHIRE	Q03376 chironomus
97	71	9.5	2738	1	PGCV_RAT	Q9erD4 rattus norv
98	70.5	9.4	335	1	G49B_MOUSE	Q64281 mus musculu
99	70.5	9.4	662	1	MM02_RABIT	P50757 oryctolagus
100	69.5	9.3	262	1	MURI_BUCAI	P57619 buchnera ap

ALIGNMENTS

RESULT 1

```
CR. HUMAN
ID   CR2_HUMAN       STANDARD;       PRT;  1033 AA.
AC   P20023.
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement receptor type 2 precursor (CR2) (Complement C3d receptor)
DE   (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).
GN   CR2 OR C4BR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91123277; PubMed=2563370;
RA   Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA   Holers V.M.;
RT   "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT   virus receptor.";
RL   J. Biol. Chem. 264:2118-2125(1989).
RN   [2]
RP   SEQUENCE OF 226-234; 256-267; 332-341; 667-677 AND 898-908.
RX   MEDLINE=86287411; PubMed=3016712;
RA   Weis J.J., Fearon D.F., Klickstein L.B., Wong W.W., Richards S.A.,
RA   de Bruyn Kops A., Smith J.A., Weis J.H.;
RT   "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT   virus receptor of human B lymphocytes: homology with the receptor for
RT   fragments C3b and C4b of the third and fourth components of
RT   complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
RN   [3]
RP   SEQUENCE OF 492-556 FROM N.A. (SHORT ISOFORM).
RX   MEDLINE=93294286; PubMed=8390533;
RA   Sinha S.K., Todd S.C., Bedrick J.A., Speiser C.L., Lambris J.D.,
RA   Tsokas C.D.;
RT   "Characterization of the EBV/C3d receptor on the human Jurkat T cell
RT   line: evidence for a novel transcript.";
RL   J. Immunol. 150:5411-5420(1993).
CC   -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN BARR
CC       VIRUS ON HUMAN B CELLS AND T CELLS. PARTICIPATES IN B LYMPHOCYTES
CC       ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES, T LYMPHOCYTES AND
CC       FOLLICULAR DENDRITIC CELLS OF THE SPLEEN.
CC   -!- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA) FAMILY.
CC   -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC   -!- DATABASE: NAME=Pro; NOTE=CD guide CD21 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd21.htm".
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; M26004; AAA35786.1; -.
DR   EMBL; M26016; AAB04638.1; -.
DR   EMBL; M24007; AAB04638.1; JOINED.
DR   EMBL; M24008; AAB04638.1; JOINED.
DR   EMBL; M24009; AAB04638.1; JOINED.
DR   EMBL; M24010; AAB04638.1; JOINED.
DR   EMBL; M24011; AAB04638.1; JOINED.
DR   EMBL; M26009; AAB04638.1; JOINED.
DR   EMBL; M26010; AAB04638.1; JOINED.
DR   EMBL; M26011; AAB04638.1; JOINED.
DR   EMBL; M26012; AAB04638.1; JOINED.
DR   EMBL; M26013; AAB04638.1; JOINED.
DR   EMBL; M26014; AAB04638.1; JOINED.
DR   EMBL; M26015; AAB04638.1; JOINED.
DR   EMBL; S62696; AAB27186.1; -.
DR   PIR; A32036; A32036.
DR   PIR; A24319; A24319.
DR   PIR; B24319; B24319.
DR   PIR; C24319; C24319.
DR   PIR; D24319; D24319.
DR   PIR; E24319; E24319.
DR   PIR; F24319; F24319.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120650; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 15.
DR   SMART; SM00032; CCP; 14.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi; Alternative splicing.
FT   SIGNAL       1     20
FT   CHAIN       21   1033       COMPLEMENT RECEPTOR TYPE 2.
FT   DOMAIN      21    971       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   972    999       POTENTIAL.
FT   DOMAIN    1000   1033       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      22     83       SUSHI 1.
FT   DOMAIN      90    147       SUSHI 2.
FT   DOMAIN     153    211       SUSHI 3.
FT   DOMAIN     214    272       SUSHI 4.
FT   DOMAIN     275    343       SUSHI 5.
FT   DOMAIN     350    407       SUSHI 6.
FT   DOMAIN     409    467       SUSHI 7.
FT   DOMAIN     470    523       SUSHI 8.
FT   DOMAIN     526    594       SUSHI 9.
FT   DOMAIN     601    658       SUSHI 10.
FT   DOMAIN     661    715       SUSHI 11.
FT   DOMAIN     718    780       SUSHI 12.
FT   DOMAIN     787    844       SUSHI 13.
FT   DOMAIN     850    908       SUSHI 14.
FT   DOMAIN     911    969       SUSHI 15.
FT   DISULFID    23     65       BY SIMILARITY.
FT   DISULFID    51     82       BY SIMILARITY.
FT   DISULFID    91    132       BY SIMILARITY.
FT   DISULFID   118    146       BY SIMILARITY.
FT   DISULFID   154    197       BY SIMILARITY.
FT   DISULFID   183    210       BY SIMILARITY.
FT   DISULFID   215    256       BY SIMILARITY.
FT   DISULFID   242    271       BY SIMILARITY.
FT   DISULFID   276    325       BY SIMILARITY.
FT   DISULFID   305    342       BY SIMILARITY.
FT   DISULFID   351    393       BY SIMILARITY.
FT   DISULFID   379    406       BY SIMILARITY.
FT   DISULFID   410    453       BY SIMILARITY.
FT   DISULFID   439    466       BY SIMILARITY.
FT   DISULFID   471    509       BY SIMILARITY.
FT   DISULFID   495    522       BY SIMILARITY.
FT   DISULFID   527    576       BY SIMILARITY.
FT   DISULFID   556    593       BY SIMILARITY.
FT   DISULFID   602    644       BY SIMILARITY.
FT   DISULFID   630    657       BY SIMILARITY.
FT   DISULFID   662    699       BY SIMILARITY.
FT   DISULFID   685    714       BY SIMILARITY.
FT   DISULFID   719    762       BY SIMILARITY.
FT   DISULFID   748    779       BY SIMILARITY.
FT   DISULFID   788    830       BY SIMILARITY.
FT   DISULFID   816    843       BY SIMILARITY.
FT   DISULFID   851    894       BY SIMILARITY.
FT   DISULFID   880    907       BY SIMILARITY.
FT   DISULFID   912    955       BY SIMILARITY.
FT   DISULFID   941    968       BY SIMILARITY.
FT   CARBOHYD   121    121       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   127    127       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   294    294       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   372    372       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   492    492       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   623    623       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   CARBOHYD     682     682       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     800     800       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     823     823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     861     861       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     911     911       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     499     524       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC     525     556       ITCPPPPPVIYNGAHTGSSLEDFPYGTTVTYTC -> NHLPT
FT                                  TPCYLQWGTHREFLRRFSIWNHGHLHM (IN SHORT
FT                                  ISOFORM).
FT   CONFLICT     667     667       Q -> D (IN REF. 2).
FT   CONFLICT     902     902       Q -> C (IN REF. 2).
FT   CONFLICT     906     906       H -> L (IN REF. 2).
SQ   SEQUENCE   1033 AA;  112973 MW;  1749DB4A07847ADA CRC64;

  Query Match             100.0%;  Score 751;  DB 1;  Length 1033;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-68;
  Matches  134;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDWTWDKPA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 GISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDWTWDKPA 79

Qy      61 PKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 PKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 139

Qy     121 PTRLPTCVSVFPLE 134
           ||||||||||||||
Db     140 PTRLPTCVSVFPLE 153


RESULT 2
CR2_MOUSE
ID   CR2_MOUSE       STANDARD;       PRT;  1025 AA.
AC   P19070;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN   CR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=90229735; PubMed=2139457;
RA   Fingeroth J.D.;
RT   "Comparative structure and evolution of murine CR2. The homolog of
RT   the human C3d/EBV receptor (CD21).";
RL   J. Immunol. 144:3458-3467(1990).
RN   [2]
RP   SEQUENCE OF 12-1025 FROM N.A.
RX   MEDLINE=91010789; PubMed=2145466;
RA   Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT   "A molecular and immunochemical characterization of mouse CR2.
RT   Evidence for a single gene model of mouse complement receptors 1 and
RT   2.";
RL   J. Immunol. 145:2974-2983(1990).
RN   [3]
RP   SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX   MEDLINE=89098890; PubMed=2783485;
RA   Fingeroth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT   "Identification of murine complement receptor type 2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN   [4]
RP   SEQUENCE OF 289-1025 FROM N.A.
RX   MEDLINE=89381350; PubMed=2528587;
RA   Kurtz C.B., Paul M.S., Aoqerter M., Weis J.J., Weis J.H.;
RT   "Murine complement receptor gene family. II. Identification and
RT   characterization of the murine homolog (Cr2) to human CR2 and its
RT   molecular linkage to Crry.";
RL   J. Immunol. 143:2058-2067(1989).
CC   -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC       LYMPHOCYTES ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC   -!- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC   -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC   ----------- ------------------------------ ------------- ------ ------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------- ------------------------------ ------------- ------ ------
DR   EMBL; M81083; AAA37451.1; -.
DR   EMBL; M35684; AAA37448.1; -.
DR   EMBL; M61132; AAA63295.1; -.
DR   EMBL; M35685; AAA37450.1; ALT_SEQ.
DR   EMBL; M29281; AAA37447.1; -.
DR   PIR; A43526; A43526.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88489; Cr2.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 14.
DR   SMART; SM00032; CCP; 14.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi.
FT   SIGNAL        1     11
FT   CHAIN        12   1025       COMPLEMENT RECEPTOR TYPE 2.
FT   DOMAIN       12    963       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    964    990       POTENTIAL.
FT   DOMAIN      991   1025       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       13     74       SUSHI 1
FT   DOMAIN       81    139       SUSHI 2
FT   DOMAIN      145    203       SUSHI 3
FT   DOMAIN      206    264       SUSHI 4
FT   DOMAIN      267    335       SUSHI 5
FT   DOMAIN      342    399       SUSHI 6
FT   DOMAIN      401    458       SUSHI 7
FT   DOMAIN      462    515       SUSHI 8
FT   DOMAIN      518    586       SUSHI 9
FT   DOMAIN      593    649       SUSHI 10.
FT   DOMAIN      653    705       SUSHI 11
FT   DOMAIN      708    769       SUSHI 12.
FT   DOMAIN      777    834       SUSHI 13.
FT   DOMAIN      840    898       SUSHI 14.
FT   DOMAIN      901    959       SUSHI 15.
FT   DISULFID     14     56       BY SIMILARITY.
FT   DISULFID     42     73       BY SIMILARITY.
FT   DISULFID     82    124       BY SIMILARITY.
FT   DISULFID    110    138       BY SIMILARITY.
FT   DISULFID    146    189       BY SIMILARITY.
FT   DISULFID    175    202       BY SIMILARITY.
FT   DISULFID    207    248       BY SIMILARITY
FT   DISULFID    234    263       BY SIMILARITY.
FT   DISULFID    268    317       BY SIMILARITY.
FT   DISULFID    297    334       BY SIMILARITY
FT   DISULFID    343    385       BY SIMILARITY.
FT   DISULFID    371    398       BY SIMILARITY.
FT   DISULFID    402    445       BY SIMILARITY.
FT   DISULFID    431    458       BY SIMILARITY.
FT   DISULFID    463    501       BY SIMILARITY.
FT   DISULFID    487    514       BY SIMILARITY.
FT   DISULFID    519    568       BY SIMILARITY.
FT   DISULFID    548    585       BY SIMILARITY.
FT   DISULFID    594    636       BY SIMILARITY.
FT   DISULFID    622    649       BY SIMILARITY.
FT   DISULFID    654    689       BY SIMILARITY.
FT   DISULFID    675    704       BY SIMILARITY
FT   DISULFID    709    752       BY SIMILARITY.
FT   DISULFID    738    769       BY SIMILARITY.
```

```
FT   DISULFID     778     820       BY SIMILARITY.
FT   DISULFID     806     838       BY SIMILARITY.
FT   DISULFID     841     884       BY SIMILARITY.
FT   DISULFID     870     89        BY SIMILARITY.
FT   DISULFID     902     945       BY SIMILARITY.
FT   DISULFID     931     959       BY SIMILARITY.
FT   CARBOHYD      77      77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     113     113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     276     276       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     316     316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     464     464       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     480     480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     484     484       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     527     527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     615     615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     639     639       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     694     694       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     754     754       N-LINKED (GLCNAC. . .) (POTENTIAL)
FT   CARBOHYD     790     790       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     813     813       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     823     823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     851     851       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     901     901       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     289     291       YGS -> EFR (IN REF. 4).
FT   CONFLICT     306     306       S -> T (IN REF. 2).
FT   CONFLICT     520     520       P -> A (IN REF. 2).
FT   CONFLICT     962     963       MISSING (IN REF. 4).
SQ   SEQUENCE   1025 AA;  112994 MW;  19E518B9A0273694 CRC64;

  Query Match          62.1%;  Score 466.5;  DB 1;  Length 1025;
  Best Local Similarity   61.2%;  Pred. No. 1.9e-39;
  Matches   82;  Conservative   29;  Mismatches   43;  Indels    2;  Gaps    1.

Qy        2 ISCCSPPPTLNGRISYYSITPIAVGTVLRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
            ||| || | | |  ||| || |||:||:|| ::|||||:: |||:::: ||||  |
Db       12 ISCTGPPEVKNARKPYYSLRYVPGTVLRYTCSPSYMIGEFAIFCLSFNQVHATWDKAPP 71

Qy       62 KCEYFNKYSSCPEPIVPGGYKIRGS-IPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
             ||   |  || ||||||:  ||| |:|||||||| || |||| |:|:|||||| |||
Db       72 ICPSVNKFIGCTSDFIVPYGIMNKSSKAFESHCDSVTFDCKANFIMEGSKIVWCQANEMWG 131

Qy      121 PTRIPTCVSVFPLE 134
            |  |  | | ||||
Db      132 PTAIPVCESDFPLE 145


RESULT 3
CR1_HUMAN
ID   CR1_HUMAN       STANDARD;       PRT;  2039 AA.
AC   P17927;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
DE   antigen).
GN   CR1 OR C3BR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88035992; PubMed=2972794;
RA   Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A.,
RA   Fearon D.T.;
RT   "Identification of distinct C3b and C4b recognition sites in the
RT   human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis.";
RL   J. Exp. Med. 168:1699-1717(1988).
RN   [2]
RP   SEQUENCE OF 503-2039 FROM N.A.
RX   MEDLINE=87168191; PubMed=2951479;
RA   Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA   Fearon D.T.;
RT   "Human C3b/C4b receptor (CR1). Demonstration of long homologous
RT   repeating domains that are composed of the short consensus repeats
RT   characteristic of C3/C4 binding proteins.";
RL   J. Exp. Med. 165:1095-1112(1987).
RN   [3]
RP   SEQUENCE OF 761-784; 831-845 AND 1179-1195 FROM N.A.
RX   MEDLINE=86067975; PubMed=2934745;
RA   Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT   "Identification of a partial cDNA clone for the human receptor for
RT   complement fragments C3b/C4b.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC   -!- FUNCTION: CR1, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC       PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC       BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC       ACTIVATED COMPLEMENT.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- POLYMORPHISM: CR1 IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.
CC   -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC       LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR-A
CC       CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC       TWO SCRS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC       SPECIFICITY.
CC   -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CR1.
CC   -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC   -!- ---------- ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---- ---- ---- ---- ---- ---- ---- ---- ---- ----
DR   EMBL; M11569; AAA52297.1; -.
DR   EMBL; M21117; AAA52298.1; -.
DR   EMBL; M21618; AAA52299.1; -.
DR   EMBL; Y00816; CAA68755.1; -.
DR   EMBL; X05409; CAA28933.1; -.
DR   PIR; A28507; A28507.
DR   PIR; A24748; A24748.
DR   PIR; B24748; B24748.
DR   PIR; C24748; C24748.
DR   PIR; S03844; S03844.
DR   HSSP; P08603; 1HF1.
DR   MIM; 120620; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi; Blood group antigen.
FT   SIGNAL       1     41
FT   CHAIN       42   2039       COMPLEMENT RECEPTOR TYPE 1.
FT   DOMAIN      42   1971       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1972   1996       POTENTIAL.
FT   DOMAIN    1997   2039       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     42     42       PYRROLIDONE CARBOXYLIC ACID (POTENTIAL)
FT   DOMAIN      42    100       SUSHI A1.
FT   DOMAIN     103    162       SUSHI A2.
FT   DOMAIN     165    233       SUSHI A3.
FT   DOMAIN     237    294       SUSHI A4.
FT   DOMAIN     296    354       SUSHI A5.
FT   DOMAIN     357    417       SUSHI A6.
FT   DOMAIN     420    488       SUSHI A7.
FT   DOMAIN     492    550       SUSHI B1.
FT   DOMAIN     553    612       SUSHI B2.
FT   DOMAIN     615    683       SUSHI B3.
FT   DOMAIN     687    744       SUSHI B4.
FT   DOMAIN     746    804       SUSHI B5.
FT   DOMAIN     807    867       SUSHI B6.
FT   DOMAIN     870    938       SUSHI B7.
FT   DOMAIN     942   1000       SUSHI C1.
FT   DOMAIN    1003   1062       SUSHI C2.
```

```
FT   DOMAIN      1065   1133       SUSHI C4.
FT   DOMAIN      1137   1194       SUSHI C4.
FT   DOMAIN      1196   1254       SUSHI C5.
FT   DOMAIN      1257   1317       SUSHI C6.
FT   DOMAIN      1320   1388       SUSHI C7.
FT   DOMAIN      1395   1453       SUSHI D1.
FT   DOMAIN      1456   1515       SUSHI D2.
FT   DOMAIN      1518   1586       SUSHI D3.
FT   DOMAIN      1590   1647       SUSHI D4.
FT   DOMAIN      1649   1707       SUSHI D5.
FT   DOMAIN      1710   1770       SUSHI D6.
FT   DOMAIN      1773   1841       SUSHI D7.
FT   DOMAIN      1847   1905       SUSHI E1.
FT   DOMAIN      1908   1966       SUSHI E2.
FT   DISULFID      43     86       BY SIMILARITY.
FT   DISULFID      73     99       BY SIMILARITY.
FT   DISULFID     104    145       BY SIMILARITY.
FT   DISULFID     131    161       BY SIMILARITY.
FT   DISULFID     166    215       BY SIMILARITY.
FT   DISULFID     195    232       BY SIMILARITY.
FT   DISULFID     248    280       BY SIMILARITY.
FT   DISULFID     265    293       BY SIMILARITY.
FT   DISULFID     297    340       BY SIMILARITY.
FT   DISULFID     326    358       BY SIMILARITY.
FT   DISULFID     358    400       BY SIMILARITY.
FT   DISULFID     386    416       BY SIMILARITY.
FT   DISULFID     421    470       BY SIMILARITY.
FT   DISULFID     450    487       BY SIMILARITY.
FT   DISULFID     493    536       BY SIMILARITY.
FT   DISULFID     523    549       BY SIMILARITY.
FT   DISULFID     554    595       BY SIMILARITY.
FT   DISULFID     581    611       BY SIMILARITY
FT   DISULFID     616    665       BY SIMILARITY.
FT   DISULFID     645    682       BY SIMILARITY.
FT   DISULFID     688    730       BY SIMILARITY.
FT   DISULFID     716    743       BY SIMILARITY.
FT   DISULFID     747    790       BY SIMILARITY.
FT   DISULFID     776    803       BY SIMILARITY.
FT   DISULFID     808    850       BY SIMILARITY.
FT   DISULFID     836    866       BY SIMILARITY.
FT   DISULFID     871    920       BY SIMILARITY.
FT   DISULFID     900    937       BY SIMILARITY.
FT   DISULFID     943    986       BY SIMILARITY.
FT   DISULFID     973    999       BY SIMILARITY.
FT   DISULFID    1004   1045       BY SIMILARITY.
FT   DISULFID    1031   1061       BY SIMILARITY.
FT   DISULFID    1066   1115       BY SIMILARITY.
FT   DISULFID    1095   1132       BY SIMILARITY.
FT   DISULFID    1138   1180       BY SIMILARITY.
FT   DISULFID    1166   1193       BY SIMILARITY.
FT   DISULFID    1197   1240       BY SIMILARITY
FT   DISULFID    1226   1258       BY SIMILARITY.
FT   DISULFID    1258   1300       BY SIMILARITY.
FT   DISULFID    1286   1316       BY SIMILARITY.
FT   DISULFID    1321   1370       BY SIMILARITY.
FT   DISULFID    1350   1387       BY SIMILARITY.
FT   DISULFID    1396   1439       BY SIMILARITY.
FT   DISULFID    1426   1452       BY SIMILARITY
FT   DISULFID    1457   1498       BY SIMILARITY.
FT   DISULFID    1484   1514       BY SIMILARITY.
FT   DISULFID    1519   1568       BY SIMILARITY.
FT   DISULFID    1548   1585       BY SIMILARITY.
FT   DISULFID    1591   1633       BY SIMILARITY.
FT   DISULFID    1619   1646       BY SIMILARITY.
FT   DISULFID    1650   1693       BY SIMILARITY.
FT   DISULFID    1679   1706       BY SIMILARITY.
FT   DISULFID    1711   1753       BY SIMILARITY.
FT   DISULFID    1749   1769       BY SIMILARITY.
FT   DISULFID    1774   1823       BY SIMILARITY.
FT   DISULFID    1803   1840       BY SIMILARITY.
FT   DISULFID    1848   1891       BY SIMILARITY.
FT   DISULFID    1877   1904       BY SIMILARITY.
FT   DISULFID    1909   1952       BY SIMILARITY.
FT   DISULFID    1938   1965       BY SIMILARITY.
FT   CARBOHYD      56     56       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     252    252       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     410    410       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     447    447       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     509    509       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     578    578       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     702    702       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     860    860       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     897    897       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     959    959       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1028   1028       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1152   1152       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1310   1310       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1481   1481       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1504   1504       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1534   1534       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1540   1540       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1605   1605       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1763   1763       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1908   1908       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   2039 AA;  223589 MW;  B2FD2336AD305EB7 CRC64;

  Query Match             29.2%;  Score 219;  DB 1;  Length 2039;
  Best Local Similarity   36.7%;  Pred. No. 1.2e-14;
  Matches   51;  Conservative   17;  Mismatches   55;  Indels   16;  Gaps    6;

Qy        2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLICITKDKV 52
            |||  || | ||   :||   |    |||: | |             | |:|:|: | :|
Db     1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVIYQC-TGPDGEQLFELVGERSIYCISKDDQ 1574

Qy       53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYYRDDSVTFACKINFSMNGNKSV 111
             | |    |:|   ||   |  | |    :: |:   :   : : | |:  | | |: ::
Db     1575 VGVWSSPPERCISINK---GIAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHIV 1631

Qy      112 WCQANNMWCPTRLPTCVSV 130
            || |  ||| :|| |  |
Db     1632 QCQTNGRWCP-KLPHCSRV 1649


RESULT 4
VCP_VACCV
ID   VCP_VACCV      STANDARD;      PRT;   243 AA.
AC   P10998;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement control protein precursor (vCP) (Secretory protein 45)
DE   (Protein C3) (28 kDa protein).
GN   C3L.
OS   Vaccinia virus (strain WR), and
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254, 10249;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-47.
RC   STRAIN=WR;
RX   MEDLINE=88318974; PubMed=3412473;
RA   Kotwal G.J., Moss B.;
RT   "Vaccinia virus encodes a secretory polypeptide structurally related
RT   to complement control proteins.";
RL   Nature 335:176-178(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WR;
RX   MEDLINE=89073756; PubMed=2849238;
RA   Kotwal G.J., Moss B.;
RT   "Analysis of a large cluster of nonessential genes deleted from a
RT   vaccinia virus terminal transposition mutant.";
RL   Virology 167:524-537(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN COPENHAGEN;
RX   MEDLINE=91121027; PubMed=2219722;
RA   Isobel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [4]
RP   COMPLETE GENOME.
RC   STRAIN COPENHAGEN;
RA   Isobel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=92115711; PubMed=1730333;
RA   Isaacs S.N., Kotwal G.J., Moss B.;
RT   "Vaccinia virus complement-control protein prevents
RT   antibody-dependent complement-enhanced neutralization of infectivity
RT   and contributes to virulence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN   [6]
RP   STRUCTURE BY NMR OF 146-260.
RX   MEDLINE=97446168; PubMed=9299352;
RA   Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
RA   Barlow P.N.;
RT   "NMR studies of a viral protein that mimics the regulators of
RT   complement activation.";
RL   J. Mol. Biol. 272:253-265(1997).
CC   -!- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
CC       INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
CC       ACTIVATION. BINDS C4B AND C4B.
CC   -!- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13176; CAA31564.1; -.
DR   EMBL; M22812; AAA69005.1; -.
DR   EMBL; M35027; AAA47997.1; -.
DR   PIR; A31005; WMVZSP.
DR   PDB; 1VVC; 03-DEC-97.
DR   PDB; 1VVD; 03-DEC-97.
DR   PDB; 1VVE; 03-DEC-97.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Signal; Repeat; Sushi; 3D-structure.
FT   SIGNAL        1     19
FT   CHAIN        20    263       COMPLEMENT CONTROL PROTEIN.
FT   DOMAIN       20     82       SUSHI 1.
FT   DOMAIN       85    144       SUSHI 2.
FT   DOMAIN      147    202       SUSHI 3.
FT   DOMAIN      205    262       SUSHI 4.
FT   DISULFID     21     70       BY SIMILARITY.
FT   DISULFID     54     81       BY SIMILARITY.
FT   DISULFID     86    126       BY SIMILARITY.
FT   DISULFID    112    143       BY SIMILARITY.
FT   DISULFID    148    190       BY SIMILARITY.
FT   DISULFID    176    201       BY SIMILARITY.
FT   DISULFID    206    248       BY SIMILARITY.
FT   DISULFID    234    261       BY SIMILARITY.
SQ   SEQUENCE   263 AA;  28029 MW;  E4322CC9A6EE8997 CRC64;

  Query Match          28.6%;  Score 215;  DB 1;  Length 263;
  Best Local Similarity   45.9%;  Pred. No. 1.1e-14;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 LSCTSPPPTINGRTSYYSITVAVGIVIFVSCCFLVCFKSIGKLFGKKSYVRLMEDAP GL
           : | ||| | ||| : |       |:| ||| : ||| :|         | | | |
Db     142 VECVSPPSISNGRHNGYEDPYTDGSVVTYSCNSGYSLIGNSGVLC     SGGEWSDP P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYRDGDSVTFACKTNFSMNGNKSVWCGANNMWG 120
            |:       || | : ||   |    | : |:| | ||  :  :|: | |   | ||
Db     200 TCQIV----KCPHPITSNGVLSSGPRRSYSYNDNVIFKCKYGYRLSGSSSSTCSPGNTWK 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 PELPKCV 262


RESULT 5
C4BP_HUMAN
ID   C4BP_HUMAN     STANDARD;      PRT:   597 AA.
AC   P04003;
DT   23-OCT-1986 (Rel. 02, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b binding protein alpha chain precursor (C4bp) (Proline rich
DE   protein) (PRP).
GN   C4BPA OR C4BP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90073699; PubMed=2590215;
RA   Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT   "Molecular cloning of the cDNA coding for proline rich protein (PRP):
RT   identity of PRP as C4b-binding protein.";
RL   Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91213199; PubMed=1989602;
RA   Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT   "Structural organization of the alpha chain of the human C4b binding
RT   protein gene.";
RL   Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN   [3]
RP   SEQUENCE OF 9-81 FROM N.A.
RX   MEDLINE=88242821; PubMed=3378624;
RA   Lintin S.J., Lewin A.R., Reid K.B.M.;
RT   "Derivation of the sequence of the signal peptide in human
RT   C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT   cDNA sequence.";
RL   FEBS Lett. 232:428-432(1988).
RN   [4]
RP   SEQUENCE OF 203-288 FROM N.A.
RX   MEDLINE=86301119; PubMed=3017751;
RA   Lintin S.J., Reid K.B.M.;
RT   "Studies on the structure of the human C4b-binding protein gene.";
RL   FEBS Lett. 204:77-81(1986).
RN   [5]
RP   SEQUENCE OF 80-597 FROM N.A.
RX   MEDLINE=86025405; PubMed=4040470;
RA   Chung L.P., Bentley D.R., Reid K.B.M.;
RT   "Molecular cloning and characterization of the cDNA coding for C4b
RT   binding protein, a regulatory protein of the classical pathway of the
RT   human complement system.";
RL   Biochem. J. 230:133-141(1985).
RN   [6]
RP   SEQUENCE OF 49-88.
RX   MEDLINE=85296001; PubMed=4033666;
RA   Chung L.P., Gagnon J., Reid K.B.M.;
RT   "Amino acid sequence studies of human C4b binding protein. N-terminal
RT   sequence analysis and alignment of the fragments produced by limited
RT   proteolysis with chymotrypsin and the peptides produced by cyanogen
```

```
RT   bromide treatment.";
RL   Mol. Immunol. 22:427-435(1985).
RN   [7]
RP   ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.
RX   MEDLINE=83221615; PubMed=6222381;
RA   Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT   "Visualization of human C4b binding protein and its complexes with
RT   vitamin K-dependent protein S and complement protein C4b.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC    -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC       AND WITH SERUM AMYLOID P COMPONENT.
CC    -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC       OF 4 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC       BETA CHAIN, A 530 kDa HOMOHEPTAMER OF ALPHA CHAINS OR A 500 kDa
CC       COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC       THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC       BINDING SITE FOR C4B AT THE END.
CC    -!- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC    -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC    -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R
CC    -!- CAUTION: IT IS UNCERTAIN WHETHER MET 1 OR MET-17 IS THE INITIATOR.
CC    ----------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to licenser@isb-sib.ch).
CC    ----------------------------------------------------------------------
DR   EMBL; M31452; AAA36507.1; -.
DR   EMBL; M62486; AAA36506.1; -.
DR   EMBL; M62475; AAA36506.1; JOINED.
DR   EMBL; M62476; AAA36506.1; JOINED.
DR   EMBL; M62477; AAA36506.1; JOINED.
DR   EMBL; M62478; AAA36506.1; JOINED.
DR   EMBL; M62479; AAA36506.1; JOINED.
DR   EMBL; M62480; AAA36506.1; JOINED.
DR   EMBL; M62481; AAA36506.1; JOINED.
DR   EMBL; M62482; AAA36506.1; JOINED.
DR   EMBL; M62484; AAA36506.1; JOINED.
DR   EMBL; M62485; AAA36506.1; JOINED.
DR   EMBL; X07853; CAA30701.1; -.
DR   EMBL; X04284; CAB51244.1; -.
DR   EMBL; X04296; CAA27839.1; -.
DR   EMBL; X02865; CAA26617.1; -.
DR   PIR; A33568; NBHUC4.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120830; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW   Polymorphism.
FT   SIGNAL       1     48
FT   CHAIN       49    597       C4B BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN      49    109       SUSHI 1.
FT   DOMAIN     112    171       SUSHI 2.
FT   DOMAIN     174    235       SUSHI 3.
FT   DOMAIN     238    295       SUSHI 4.
FT   DOMAIN     298    361       SUSHI 5.
FT   DOMAIN     364    423       SUSHI 6.
FT   DOMAIN     425    481       SUSHI 7.
FT   DOMAIN     483    539       SUSHI 8.
FT   DISULFID    50     96       BY SIMILARITY.
FT   DISULFID    81    108       BY SIMILARITY.
FT   DISULFID   114    154       BY SIMILARITY.
FT   DISULFID   140    170       BY SIMILARITY.
FT   DISULFID   175    217       BY SIMILARITY.
FT   DISULFID   203    234       BY SIMILARITY.
FT   DISULFID   239    281       BY SIMILARITY.
FT   DISULFID   267    294       BY SIMILARITY.
FT   DISULFID   299    348       BY SIMILARITY.
FT   DISULFID   332    360       BY SIMILARITY.
FT   DISULFID  ?364   ?387       BY SIMILARITY.
FT   DISULFID  ?365   ?409       BY SIMILARITY.
FT   DISULFID  ?399    422       BY SIMILARITY.
FT   DISULFID   426    468       BY SIMILARITY.
FT   DISULFID   454    480       BY SIMILARITY.
FT   DISULFID   484    525       BY SIMILARITY.
FT   DISULFID   511    538       BY SIMILARITY.
FT   DISULFID   546    546       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID   558    558       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   CARBOHYD   221    221       N-LINKED (GLCNAC...).
FT   CARBOHYD   506    506       N-LINKED (GLCNAC...).
FT   CARBOHYD   528    528       N-LINKED (GLCNAC...).
FT   VARIANT     92     92       Q -> T.
FT                               /FTId=VAR_001977.
FT   VARIANT    357    357       Y -> S.
FT                               /FTId=VAR_001978.
FT   VARIANT    473    473       W -> * (in DBSNP:1801341).
FT                               /FTId=VAR_012038.
SQ   SEQUENCE   597 AA;  67033 MW;  67E3F2AB5A16DD CRC64;

  Query Match             26.5%;  Score 199;  DB 1;  Length 597;
  Best Local Similarity   32.3%;  Pred. No. 1.2e-12;
  Matches   41;  Conservative   20;  Mismatches   60;  Indels    6;  Gaps    3;

Qy       2 ISCGSPPPEILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           : |  || | ||| |       | |  : |||   | |:|  |: |  ::  | |    |
Db     173 VKCKPPPEIRNGRHSGEENEYAVGESVEYSCDPRFSLIGHASISCIVENETIGVWRPSPP 232

Qy      62 KCEYENKYSSCPEPIVPGGYKIRGSTP-YRHGSVIEFACKTNFSMNGNKSVWCQANNMWG 120
           ||      :| :| |  |  : |  | | : |:: | |:  | : |:  : | |:: |
Db     233 TCEKI----TCRKPDVSHGEMVSGFGPIYNYKATIVFKCKEGFVLRGSSVIHCDADSKWN 288

Qy     121 PTRLPTC 127
           |:  | |
Db     289 PSP-PAC 294


RESULT 6
C4BP_BOVIN
ID   C4BP_BOVIN     STANDARD;      PRT:   610 AA.
AC   Q28065;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein alpha chain precursor (C4bp).
GN   C4BPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver.
RX   MEDLINE=95015909; PubMed=7930621;
RA   Hillarp A., Thern A., Dahlbaech B.;
RT   "Bovine C4b binding protein. Molecular cloning of the alpha- and
RT   beta chains provides structural background for lack of complex
RT   formation with protein S.";
RL   J. Immunol. 153:4190-4199(1994).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
```

```
CC       (BY SIMILARITY).
CC    -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC    -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC       -------- -- ---- - ---- ---- ---- ---- ---- ---- ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC       -------- -- ---- - ---- ---- ---- ---- ---- ---- ---- ----
DR   EMBL; Z31693; CAA86498.1;  -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   "Complement" pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1     48       BY SIMILARITY.
FT   CHAIN        49    610       C4B-BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN       49    108       SUSHI 1.
FT   DOMAIN      111    170       SUSHI 2.
FT   DOMAIN      173    235       SUSHI 3.
FT   DOMAIN      238    295       SUSHI 4.
FT   DOMAIN      298    363       SUSHI 5.
FT   DOMAIN      366    426       SUSHI 6.
FT   DOMAIN      428    484       SUSHI 7.
FT   DOMAIN      486    542       SUSHI 8.
FT   DISULFID     50     95       BY SIMILARITY.
FT   DISULFID     80    107       BY SIMILARITY.
FT   DISULFID    112    153       BY SIMILARITY.
FT   DISULFID    139    169       BY SIMILARITY.
FT   DISULFID    174    217       BY SIMILARITY.
FT   DISULFID    203    234       BY SIMILARITY.
FT   DISULFID    239    281       BY SIMILARITY.
FT   DISULFID    267    294       BY SIMILARITY.
FT   DISULFID    299    350       BY SIMILARITY.
FT   DISULFID    344    362       BY SIMILARITY.
FT   DISULFID   2366   2390       BY SIMILARITY.
FT   DISULFID   2367   2412       BY SIMILARITY.
FT   DISULFID   2402    429       BY SIMILARITY.
FT   DISULFID    429    471       BY SIMILARITY
FT   DISULFID    457    483       BY SIMILARITY.
FT   DISULFID    487    528       BY SIMILARITY.
FT   DISULFID    514    541       BY SIMILARITY.
FT   DISULFID    549    549       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID    561    561       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   CARBOHYD     66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    221    221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    525    525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    602    602       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   610 AA;  68846 MW;  DB06B270EBA06B58 CRC64;

  Query Match             26.9%;  Score 197.5;  DB 1;  Length 610;
  Best Local Similarity   42.9%;  Pred. No. 1.7e-12;
  Matches   42;  Conservative   21;  Mismatches   52;  Indels   15;  Gaps    5;

Qy       4 CGSPPPTTNGRIS      YYSTTAVGTVTKYSCSGTFRLIGEKSLICITKDKVDGTWDK 58
           |  || | ||  :       ||     |: : |||   | :|: |: |  ::|  | |
Db     174 CEPPPTISNGRHNGGDEDPYI    YGSSVTYSCDRDFSMLGKASISCRVENKTIGVWSP 229

Qy      59 PAPKCEYFNKYSS-PEPIVEGYKIRYSIP YRIPDSVTEACKINFSMNGNFSVWQANN 117
           | |    |  | :|:   |    | | | :  |: :||   | : |:  : |:|:|
Db     230 SPPSC    KKVI VQPVVEPCKLPSCEGPITYYQQSIVYACNKGFRLEGDNLIHCTADN 285

Qy     118 MWGPTRLPTC 127
           | |   |||
Db     286 SWNPEP PIC 294


RESULT 7
C4BP_RAT
ID   C4BP_RAT       STANDARD;      PRT:   558 AA.
AC   Q63514;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein alpha chain precursor (C4bp).
GN   C4BPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=97166082; PubMed=9013975;
RA   Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT   "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
RT   structural and functional relationships among human, bovine, rabbit
RT   mouse, and rat proteins.";
RL   J. Immunol. 158:1315-1323(1997).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4B*2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC       AND WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC       -------- -- ---- - ---- ---- ---- ---- ---- ---- ---- ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC       -------- -- ---- - ---- ---- ---- ---- ---- ---- ---- ----
DR   EMBL; Z50051; CAA90391.1;  -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1     13       BY SIMILARITY.
FT   CHAIN        14    558       C4B-BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN       14     73       SUSHI 1.
FT   DOMAIN       76    135       SUSHI 2.
FT   DOMAIN      138    200       SUSHI 3.
FT   DOMAIN      203    259       SUSHI 4.
FT   DOMAIN      262    325       SUSHI 5.
FT   DOMAIN      328    387       SUSHI 6.
FT   DOMAIN      389    444       SUSHI 7.
FT   DOMAIN      446    502       SUSHI 8.
FT   DISULFID     15     60       BY SIMILARITY.
FT   DISULFID     45     72       BY SIMILARITY.
FT   DISULFID     77    118       BY SIMILARITY.
FT   DISULFID    104    134       BY SIMILARITY.
FT   DISULFID    139    182       BY SIMILARITY.
FT   DISULFID    168    199       BY SIMILARITY.
FT   DISULFID    204    246       BY SIMILARITY.
FT   DISULFID    232    258       BY SIMILARITY.
FT   DISULFID    263    312       BY SIMILARITY.
FT   DISULFID    296    324       BY SIMILARITY.
FT   DISULFID   2328   2351       BY SIMILARITY.
FT   DISULFID   2329   2373       BY SIMILARITY.
FT   DISULFID   2364    386       BY SIMILARITY.
FT   DISULFID    390    431       BY SIMILARITY.
FT   DISULFID    417    443       BY SIMILARITY.
FT   DISULFID    447    488       BY SIMILARITY.
FT   DISULFID    474    501       BY SIMILARITY.
FT   DISULFID    509    509       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID    521    521       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
```

```
FT   CARBOHYD      81      81       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     177     177       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     186     186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     469     469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     491     491       N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ   SEQUENCE   558 AA;  62266 MW;  592F0C667ED1E5FF CRC64;

  Query Match             24.2%;  Score 182;  DB 1;  Length 558;
  Best Local Similarity   29.9%;  Pred. No. 5.6e-11;
  Matches   41;  Conservative   18;  Mismatches   48;  Indels   30;  Gaps    6;

Qy       4 CGSPPPILNGR--------ISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGT 55
           || || | ||:          :| |:        : | |   | |:|  |: |   :|  |
Db     139 CGMPPDISNGKHNGREEEEFTYRSS-------VTYKCDPDFTLLGNASITCTVVNKTVGV 191

Qy      56 WDKPAPKCEYFNKYSSCPEP-   - IVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKS 110
           |    | ||        || |       :  |:|      |:: ||| | |:  | : |:
Db     192 WSPSPPTCERI----ICPWPKVLRGEINSGFK----HTYKYKDSVRFVCQKGFVLRGSGV 243

Qy     111 VWCQANNMWGPTRLPTC 127
           : |:|:  | |  :| |
Db     244 IHCEADGSWSP--VPVC 258


RESULT 8
CFAH_HUMAN
ID   CFAH_HUMAN      STANDARD;      PRT;  1231 AA.
AC   P08603; Q14570; P78435; Q9NU86;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor H precursor (H factor 1).
GN   HF1 OR HF OR CFH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC   TISSUE=Liver;
RX   MEDLINE=88134059; PubMed=2963625;
RA   Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;
RT   "The complete amino acid sequence of human complement factor H ";
RL   Biochem. J. 249:593-602(1988).
RN   [2]
RP   SEQUENCE OF 53-445 FROM N.A.
RX   MEDLINE=87054207; PubMed=2946589;
RA   Schulz T.F., Schwaeble W., Stanley K.K., Weiss E., Dierich M.P.;
RT   "Human complement factor H: isolation of cDNA clones and partial cDNA
RT   sequence of the 48 kDa tryptic fragment containing the binding site
RT   for C4b.";
RL   Eur. J. Immunol. 16:1351-1355(1986).
RN   [3]
RP   SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=86169701; PubMed=2937845;
RA   Kristensen T., Wetsel R.A., Tack B.F.;
RT   "Structural analysis of human complement protein H: homology with C4b
RT   binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
RL   J. Immunol. 136:3407-3411(1986).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Bird C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1047-1231 FROM N.A.
RX   MEDLINE=91201892; PubMed=1826708;
RA   Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;
RT   "Cloning of the 1.4-kb mRNA species of human complement factor H
RT   reveals a novel member of the short consensus repeat family related
RT   to the carboxy terminal of the classical 150-kDa molecule.";
RL   J. Immunol. 146:3190-3196(1991).
RN   [6]
RP   SEQUENCE OF 19-45.
RX   MEDLINE=83048213; PubMed=6215918;
RA   Sim R.B., Discipio R.G.;
RT   "Purification and structural studies on the complement-system control
RT   protein beta 1H (Factor H).";
RL   Biochem. J. 209:285-293(1982).
RN   [7]
RP   SEQUENCE OF 1-19 FROM N.A.
RA   Vik D.P., Williams S.A.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 1-9 FROM N.A.
RA   Dominguez O.;
RL   Thesis (1993), Hospital Trias I Pujo  Spain.
RN   [9]
RP   STRUCTURE BY NMR OF 927-985 (SUSHI 16).
RX   MEDLINE=9 278097; PubMed=1829116;
RA   Norman D.G., Barlow P.N., Baron M., Day A.J., Sim R., Campbell I.D.,
RT   "Three-dimensional structure of a complement control protein module
RT   in solution.";
RL   J. Mol. Biol. 219:717-725(1991).
RN   [10]
RP   STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX   MEDLINE=92252649; PubMed=1533152;
RA   Barlow P.N., Norman D.G., Steinkasserer A., Horne I.J., Pearce J.,
RA   Driscoll P.C., Sim R., Campbell I.D.;
RT   "Solution structure of the fifth repeat of factor H: a second example
RT   of the complement control protein module.";
RL   Biochemistry 31:3626-3634(1992).
RN   [11]
RP   STRUCTURE BY NMR OF 866-985 (SUSHIS 15 AND 16).
RX   MEDLINE=93323119; PubMed=8331663;
RA   Barlow P.N.  Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RA   Sim R., Campbell I.D.;
RT   "Solution structure of a pair of complement modules by nuclear
RT   magnetic resonance.";
RL   J. Mol. Biol. 232:268-284(1993).
CC   -!- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC       C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC       C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC       CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS  1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC       341 ONWARD DUE TO A FRAMESHIFT.
CC   ---------- ---------- ---  ----------  ---  ----------  ---  --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---  ----------  ---  ----------  ---  --
DR   EMBL; Y00716; CAA68704.1; -.
DR   EMBL; X04697; CAH41739.1; ALT_FRAME.
DR   EMBL; X07523; CAA30403.1; -.
DR   EMBL; M12383; AAA52013.1; -.
DR   EMBL; AL049744; CAB70598.1; -.
DR   EMBL; M65294; AAA35948.1; -.
DR   EMBL; U56979; AAH01987.1; -.
DR   EMBL; Z29665; CAA82763.1; -.
DR   PIR; S00251; NBHUH.
DR   PIR; S03013; S03013.
DR   PDB; 1HCC; 15-APR-92.
DR   PDB; 1HFH; 15-JUL-93.
DR   PDB; 1HFI; 15-JUL-93.
DR   MIM; 134370; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP
DR   Pfam; PF00084; sushi; 20.
DR   SMART; SM00032; CCP; 20.
KW   Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW   Signal; 3D-structure; Polymorphism  Alternative splicing.
```

```
FT   SIGNAL        1     18
FT   CHAIN        19   1231       COMPLEMENT FACTOR B.
FT   DOMAIN       20     81       SUSHI 1.
FT   DOMAIN       84    142       SUSHI 2.
FT   DOMAIN      145    206       SUSHI 3.
FT   DOMAIN      209    263       SUSHI 4.
FT   DOMAIN      266    321       SUSHI 5.
FT   DOMAIN      424    386       SUSHI 6.
FT   DOMAIN      488    443       SUSHI 7.
FT   DOMAIN      447    506       SUSHI 8.
FT   DOMAIN      508    565       SUSHI 9.
FT   DOMAIN      568    624       SUSHI 10.
FT   DOMAIN      629    685       SUSHI 11.
FT   DOMAIN      690    745       SUSHI 12.
FT   DOMAIN      752    804       SUSHI 13.
FT   DOMAIN      810    865       SUSHI 14.
FT   DOMAIN      869    917       SUSHI 15.
FT   DOMAIN      930    985       SUSHI 16.
FT   DOMAIN      988   1044       SUSHI 17.
FT   DOMAIN     1047   1103       SUSHI 18.
FT   DOMAIN     1108   1164       SUSHI 19.
FT   DOMAIN     1166   1229       SUSHI 20.
FT   DISULFID     21     66       BY SIMILARITY.
FT   DISULFID     52     80       BY SIMILARITY.
FT   DISULFID     85    129       BY SIMILARITY.
FT   DISULFID    114    141       BY SIMILARITY.
FT   DISULFID    146    192       BY SIMILARITY.
FT   DISULFID    178    205       BY SIMILARITY.
FT   DISULFID    210    251       BY SIMILARITY.
FT   DISULFID    247    262       BY SIMILARITY.
FT   DISULFID    267    309       BY SIMILARITY.
FT   DISULFID    294    320       BY SIMILARITY.
FT   DISULFID    425    374       BY SIMILARITY.
FT   DISULFID    357    385       BY SIMILARITY.
FT   DISULFID    389    432       BY SIMILARITY.
FT   DISULFID    416    442       BY SIMILARITY.
FT   DISULFID    448    493       BY SIMILARITY.
FT   DISULFID    477    505       BY SIMILARITY.
FT   DISULFID    509    555       BY SIMILARITY.
FT   DISULFID    546    564       BY SIMILARITY.
FT   DISULFID    569    611       BY SIMILARITY.
FT   DISULFID    597    625       BY SIMILARITY.
FT   DISULFID    630    675       BY SIMILARITY.
FT   DISULFID    659    684       BY SIMILARITY.
FT   DISULFID    691    733       BY SIMILARITY.
FT   DISULFID    739    744       BY SIMILARITY.
FT   DISULFID    753    792       BY SIMILARITY.
FT   DISULFID    781    803       BY SIMILARITY.
FT   DISULFID    811    853       BY SIMILARITY.
FT   DISULFID    839    864       BY SIMILARITY.
FT   DISULFID    870    915       BY SIMILARITY.
FT   DISULFID    901    926       BY SIMILARITY.
FT   DISULFID    941    973       BY SIMILARITY.
FT   DISULFID    959    984       BY SIMILARITY.
FT   DISULFID    989   1032       BY SIMILARITY.
FT   DISULFID   1018   1043       BY SIMILARITY.
FT   DISULFID   1048   1091       BY SIMILARITY.
FT   DISULFID   1077   1102       BY SIMILARITY.
FT   DISULFID   1109   1152       BY SIMILARITY.
FT   DISULFID   1138   1163       BY SIMILARITY.
FT   DISULFID   1167   1218       BY SIMILARITY.
FT   DISULFID   1201   1228       BY SIMILARITY.
FT   CARBOHYD    529    529       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    718    718       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    802    802       N-LINKED (GLCNAC...).
FT   CARBOHYD    822    822       N-LINKED (GLCNAC...).
FT   CARBOHYD    882    882       N-LINKED (GLCNAC...).
FT   CARBOHYD    911    911       N-LINKED (GLCNAC...).
FT   CARBOHYD   1029   1029       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1095   1095       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    446    449       KTCS -> SFTL (IN ISOFORM 2).
FT   VARSPLIC    450   1231       MISSING (IN ISOFORM 2).
FT   VARIANT     402    402       H -> Y.
FT                                 /FTId=VAR_001979.
FT   CONFLICT     21     21       C -> Q (IN REF. 3).
FT   CONFLICT     30     30       T -> V (IN REF. 3).
FT   CONFLICT     34     34       I -> Q (IN REF. 3).
FT   CONFLICT     53     54       RP -> IL (IN REF. 2).
FT   STRAND      870    871
FT   STRAND      876    876

  Query Match             24.0%;  Score 180.5;  DB 1;  Length 1231;
  Best Local Similarity   34.8%;  Pred. No. 1.9e-10;
  Matches   47;  Conservative   24;  Mismatches   50;  Indels   15;  Gaps    8;

Qy        3 SCCSPPPEINGRISYYSTDIAVG  EVEYKCNSGFELVGEKSLICIKKDKVEWDRPKP 60
            :|| :| :||| :   |    |   |: | | |:  | : :   :|     ||| |
Db      629 SCGPPPELLNGNVK-EKTKEEYGHSEVVEYYCNPRFLMKGPNKIQC-    VDGEW TTL 681

Qy       61 PKCEYPNEYSSGIG-DIVFKDRYFRSSIDFRGTEVISVEIAPENESMNYNEAVWQANNMW 119
             | |    : |:| : | :  |:    | || :|||| | |  :::: :::|: |  : ::
Db      682 PVC  IVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCSESFTMIGHRSIIC HGVW TGE 738

Qy      120 GPTKLPTCVSVFPLE 134
              :|| |||:  |:
Db      739   TQLPQCVAIDKLK 751


RESULT 9
C4BP_MOUSE
ID   C4BP_MOUSE     STANDARD;      PRT;   469 AA.
AC   P08607;
DT   01-AUG 1988 (Rel. 08, Created)
DT   01-FEB 1994 (Rel. 28, Last sequence update)
DT   16-OCT 2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein precursor (C4bp).
GN   C4BPA OR C4BP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88024997; PubMed=3663616;
RA   Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT   "cDNA structure of murine C4b-binding protein, a regulatory component
RT   of the serum complement system.";
RL   Biochemistry 26:4668-4674(1987).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3biNA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: HOMOHEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
CC       BETA CHAIN OF C4BP.
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET 1 OR MET 44 IS THE INITIATOR.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M17122; AAA37312.1; ALT_INIT.
DR   PIR; A27117; NBMSC4.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88229; C4bp.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
```

```
FT   SIGNAL        1     56
FT   CHAIN        57    469       C4B BINDING PROTEIN.
FT   DOMAIN       57    116       SUSHI 1.
FT   DOMAIN      119    177       SUSHI 2.
FT   DOMAIN      180    241       SUSHI 3.
FT   DOMAIN      244    300       SUSHI 4.
FT   DOMAIN      302    356       SUSHI 5.
FT   DOMAIN      358    414       SUSHI 6.
FT   DISULFID     58    103       BY SIMILARITY.
FT   DISULFID     88    115       BY SIMILARITY.
FT   DISULFID    120    160       BY SIMILARITY.
FT   DISULFID    146    176       BY SIMILARITY.
FT   DISULFID    181    223       BY SIMILARITY.
FT   DISULFID    209    240       BY SIMILARITY.
FT   DISULFID    245    287       BY SIMILARITY.
FT   DISULFID    273    299       BY SIMILARITY.
FT   DISULFID    303    343       BY SIMILARITY.
FT   DISULFID    329    355       BY SIMILARITY.
FT   DISULFID    359    400       BY SIMILARITY.
FT   DISULFID    386    413       BY SIMILARITY.
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    227    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    275    275       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    366    366       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    428    428       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   469 AA;  51551 MW;  41E137CB8D8C6321 CRC64;

  Query Match             23.8%;  Score 178.5;  DB 1;  Length 469;
  Best Local Similarity   33.3%;  Pred. No. 1e-10;
  Matches   45;  Conservative   13;  Mismatches   54;  Indels   23;  Gaps    7;

Qy        2 ISCGSPPPIINGRIS----YYSIITAVGIVIRYSCSGTERLIGEKSLILCITKDKVDGTWD 57
            : || || | ||: |    :|  |   |  | |:|   |||:|   : |  :|    |
Db      179 VKCGPPPDISNGKHSGTEDFY--PYNEG--ISYTCDEGFRLVGSPFIGCIVVNKIVPVWS 234

Qy       58 KPAPKCEYFNKYSSCPPP-- --  IVPGGYKIRGSTPYPHGDSVTFACKTNFSMNGNKSVW 112
               | ||        | :|     ::  |||      | | |||  ||     : |   :
Db      235 SSPPTCEKI     ICSQPNIIHGVIVSGYK----ATYTHRDSVRLACLNGTVLRGRHVIE 286

Qy      113 CQANNMWGPTRLPTC 127
            || |  |  : ||||
Db      287 CQGNGNW--SSLPTC 299


RESULT 10
CFAH_MOUSE
ID   CFAH_MOUSE     STANDARD;      PRT:  1234 AA.
AC   P06909;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement factor H precursor (Protein beta-1-H).
GN   HF1 OR CFH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86243353; PubMed=2940596;
RA   Kristensen T., Tack B.F.;
RT   "Murine protein H is comprised of 20 repeating units, 61 amino acids
RT   in length.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN   [2]
RP   SEQUENCE OF 1-19 FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=90148935; PubMed=2533512;
RA   Munoz-Canoves P., Tack B.F., Vik D.P.;
RT   "Analysis of complement factor H mRNA expression: dexamethasone and
RT   IFN-gamma increase the level of H in L cells.";
RL   Biochemistry 28:9891-9897(1989).
RN   [3]
RP   SEQUENCE OF 1-18 FROM N.A.
RX   MEDLINE=90111033; PubMed=2136885;
RA   Natsuume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C.,
RA   Moriwaki K.;
RT   "Demonstration of an unusual allelic variation of mouse factor H by
RT   the complete cDNA sequence of the H-2 allotype.";
RL   J. Immunol. 144:358-362(1990).
CC   -!- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC       C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC       C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC       CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC   -!- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
CC       MICE.
CC   -!- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12660; AAA37759.1; -.
DR   EMBL; J02891; AAA37795.1; -.
DR   EMBL; M31979; AAA37762.1; -.
DR   PIR; A26154; NBMSH.
DR   HSSP; P08603; 1HFI.
DR   MGD; MGI:88385; Cfh.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 20.
DR   SMART; SM00032; CCP; 20.
KW   Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW   Signal.
FT   SIGNAL        1     18       BY SIMILARITY.
FT   CHAIN        19   1234       COMPLEMENT FACTOR H.
FT   DOMAIN       20     81       SUSHI 1.
FT   DOMAIN       84    142       SUSHI 2.
FT   DOMAIN      145    206       SUSHI 3.
FT   DOMAIN      209    263       SUSHI 4.
FT   DOMAIN      266    321       SUSHI 5.
FT   DOMAIN      324    386       SUSHI 6.
FT   DOMAIN      388    443       SUSHI 7.
FT   DOMAIN      447    506       SUSHI 8.
FT   DOMAIN      508    565       SUSHI 9.
FT   DOMAIN      568    623       SUSHI 10.
FT   DOMAIN      628    684       SUSHI 11.
FT   DOMAIN      689    744       SUSHI 12.
FT   DOMAIN      751    803       SUSHI 13.
FT   DOMAIN      807    862       SUSHI 14.
FT   DOMAIN      866    932       SUSHI 15.
FT   DOMAIN      935    990       SUSHI 16.
FT   DOMAIN      993   1049       SUSHI 17.
FT   DOMAIN     1052   1108       SUSHI 18.
FT   DOMAIN     1113   1169       SUSHI 19.
FT   DOMAIN     1171   1234       SUSHI 20.
FT   DISULFID     21     66       BY SIMILARITY.
FT   DISULFID     52     80       BY SIMILARITY.
FT   DISULFID     85    129       BY SIMILARITY.
FT   DISULFID    114    141       BY SIMILARITY.
FT   DISULFID    146    192       BY SIMILARITY.
FT   DISULFID    173    205       BY SIMILARITY.
FT   DISULFID    210    251       BY SIMILARITY.
FT   DISULFID    237    262       BY SIMILARITY.
FT   DISULFID    267    309       BY SIMILARITY.
FT   DISULFID    294    320       BY SIMILARITY.
FT   DISULFID    325    374       BY SIMILARITY.
FT   DISULFID    357    385       BY SIMILARITY.
FT   DISULFID    389    431       BY SIMILARITY.
FT   DISULFID    416    442       BY SIMILARITY.
FT   DISULFID    440    494       BY SIMILARITY.
```

```
FT   DISULFID    477    505       BY SIMILARITY.
FT   DISULFID    509    553       BY SIMILARITY.
FT   DISULFID    536    564       BY SIMILARITY.
FT   DISULFID    569    610       BY SIMILARITY.
FT   DISULFID    597    622       BY SIMILARITY.
FT   DISULFID    629    672       BY SIMILARITY.
FT   DISULFID    658    683       BY SIMILARITY.
FT   DISULFID    690    732       BY SIMILARITY.
FT   DISULFID    718    743       BY SIMILARITY.
FT   DISULFID    752    791       BY SIMILARITY.
FT   DISULFID    780    802       BY SIMILARITY.
FT   DISULFID    808    850       BY SIMILARITY.
FT   DISULFID    836    861       BY SIMILARITY.
FT   DISULFID    867    920       BY SIMILARITY.
FT   DISULFID    906    931       BY SIMILARITY.
FT   DISULFID    936    978       BY SIMILARITY.
FT   DISULFID    964    989       BY SIMILARITY.
FT   DISULFID    994   1037       BY SIMILARITY.
FT   DISULFID   1023   1048       BY SIMILARITY.
FT   DISULFID   1054   1096       BY SIMILARITY.
FT   DISULFID   1082   1107       BY SIMILARITY.
FT   DISULFID   1114   1152       BY SIMILARITY.
FT   DISULFID   1143   1168       BY SIMILARITY.
FT   DISULFID   1172   1223       BY SIMILARITY.
FT   DISULFID   1206   1233       BY SIMILARITY.
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    773    773       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    801    801       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1030   1030       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1225   1225       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1234 AA;  136082 MW;  C5AC02F341B957F7 CRC64;

  Query Match             23.8%;  Score 178.5;  DB 1;  Length 1234;
  Best Local Similarity   37.0%;  Pred. No. 3e-10;
  Matches   51;  Conservative   14;  Mismatches   52;  Indels   21;  Gaps    8;

Qy        4 SCGSPPPILNGRISYYSITYAVGTIVKYSCSGTFRLIGEKSLLCITKDKVDGTWDKP 63
            ||  |  ||| |:      ||     | |::| |   | | .  : |     ||| |
Db      628 SCAPPLELLNGEINGAKKVEYSH    GEVVKYICKPGFELEGPNKIQC---  VDGNW- 677

Qy       64 KPAPKCEYFNKYSSCPE PIVPGGYKIRGSTPYRHGDSVTFACETNFSMNGNKSVWCQAN 123
               | |    :  :|  : |  :  |       || ||||| |  : ||:| :: || |  :
Db      678 LTLPVC  TEEERICGDDIPELEHGSAKCSVPPYHHGDSVEFTCEENFTMIGHCGVSC 1S 734

Qy      124 NMWGPTRLPTCVSVFPLE 141
              |  :||| ||:  |:
Db      735 GKW  IQLPKCVAIDQLE 752

RESULT 11
DAF1_MOUSE
ID   DAF1_MOUSE     STANDARD;      PRT:   390 AA.
AC   Q61475; Q63497; P97742;
DT   31-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Complement decay-acceleration factor, GPI-anchored precursor
DE   (DAF GPI).
GN   DAF1
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=95403982; PubMed=7545711;
RA   Spicer A.P., Seldin M.F., Gendler S.J.;
RT   "Molecular cloning and chromosomal localization of the mouse decay
RT   acceleration factor genes. Duplicated genes encode
RT   glycosylphosphatidylinositol-anchored and transmembrane forms.";
RL   J. Immunol. 155:3079-3091(1995).
RN   [2]
RP   SEQUENCE OF 7-390 FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Spleen;
RX   MEDLINE=96462213; PubMed=8871624;
RA   Fukuoka Y., Yasui A., Okada N., Okada H.;
RT   "Molecular cloning of murine decay accelerating factor by
RT   immunoscreening.";
RL   Int. Immunol. 8:879-885(1996).
CC   -!- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor.
CC   -!- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,
CC       LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC   -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L41366; AAB00091.1; -.
DR   EMBL; D63679; BAA09830.1; -.
DR   HSSP; P08603; 1HCC.
DR   MGD; MGI:104856; Daf1.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;
KW   Signal; Sushi.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    361       COMPLEMENT DECAY-ACCELERATING FACTOR,
FT                                GPI-ANCHORED.
FT   PROPEP      362    390       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       35     95       SUSHI 1.
FT   DOMAIN       97    159       SUSHI 2.
FT   DOMAIN      162    221       SUSHI 3.
FT   DOMAIN      224    285       SUSHI 4.
FT   DOMAIN      288    364       SER/THR RICH.
FT   DISULFID     65     94       BY SIMILARITY.
FT   DISULFID     98    145       BY SIMILARITY.
FT   DISULFID    129    158       BY SIMILARITY.
FT   DISULFID    164    204       BY SIMILARITY.
FT   DISULFID    190    220       BY SIMILARITY.
FT   DISULFID    225    267       BY SIMILARITY.
FT   DISULFID    254    284       BY SIMILARITY.
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       361    361       GPI-ANCHOR (BY SIMILARITY).
FT   CONFLICT      7      7       P  - A (IN REF. 2).
FT   CONFLICT      9      9       T  - A (IN REF. 2).
FT   CONFLICT     83     83       E  - G (IN REF. 2).
FT   CONFLICT     91     91       E  - G (IN REF. 2).
FT   CONFLICT    135    135       E  - K (IN REF. 2).
FT   CONFLICT    173    173       H  - L (IN REF. 2).
FT   CONFLICT    180    180       I  - T (IN REF. 2).
SQ   SEQUENCE   390 AA;  42618 MW;  44187210EF47E8E7 CRC64;

  Query Match             22.3%;  Score 167.5;  DB 1;  Length 390;
  Best Local Similarity   31.7%;  Pred. No. 1.1e-09;
  Matches   45;  Conservative   20;  Mismatches   52;  Indels   2%;  Gaps    7;

Qy        3 SCGSPPPILNGRISYYSITYAVGTIVKYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 62
            ||  ::  : || |:    | :  |: | :||:  :|| |  |  | | |:  :  |
Db      162 SCPNPKDLDNGHIN-IPTGILEGSEINFSCNPGYRLVGVSSTFCSVIGNIVD  WEDEFP 218
```

```
Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTPYYRGDSVTFACKTNFSMNGNKSVWCQAN---- 116
           |   :    ||||       :|| |  | :   ||:::|   | : || |:::|  :
Db     219 VCTEIR--GCPEPPKINNGIMRGESDSYTYSQVVTYSCDKGFILVGNASIYCTVSKSDV 274

Qy     117 NMWG------PTRLPT 126
           |          ||: ||
Db     275 GQWSSPPPRCIEKSKVPTKKPT 296


RESULT 12
F13B_HUMAN
ID   F13B_HUMAN      STANDARD;      PRT;   661 AA.
AC   P05160;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein-
DE   glutamine gamma-glutamyltransferase B chain) (Transglutaminase B
DE   chain).
GN   F13B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91105054; PubMed=2271707;
RA   Bottenus R.E., Ichinose A., Davie E.W.;
RT   "Nucleotide sequence of the gene for the b subunit of human factor
RT   XIII.";
RL   Biochemistry 29:11195-11209(1990).
RN   [2]
RP   SEQUENCE OF 2-661 FROM N.A.
RX   MEDLINE=87026535; PubMed=3021194;
RA   Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;
RT   "Amino acid sequence of the b subunit of human factor XIII, a protein
RT   composed of ten repetitive segments.";
RL   Biochemistry 25:4633-4638(1986).
RN   [3]
RP   REVISIONS.
RA   Ichinose A.;
RL   Submitted (FEB-1987) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-20 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90251467; PubMed=2339067;
RA   Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;
RT   "Complete cDNA sequence encoding the B subunit of human factor XIII.";
RL   Nucleic Acids Res. 18:2817-2817(1990).
RN   [5]
RP   VARIANT PHE 450.
RX   MEDLINE=93313189; PubMed=8324218;
RA   Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;
RT   "Two genetic defects in a patient with complete deficiency of the b-
RT   subunit for coagulation factor XIII.";
RL   Blood 82:145-150(1993).
CC   -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC   -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS
CC   -!- DISEASE: A DEFICIENCY IN FA13 CAN RESULT IN A LIFELONG BLEEDING
CC       TENDENCY, DEFECTIVE WOUND HEALING, AND HABITUAL ABORTION.
CC   -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M64554; AAA51821.1; ALT_SEQ.
DR   EMBL; M14057; AAA88042.1; -.
DR   EMBL; X51823; CAA36123.1; -.
DR   PIR; A23830; A23830.
DR   PIR; A36397; A36397.
DR   PIR; S09980; S09980.
DR   HSSP; P08603; 1HFI.
DR   MIM; 134580; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi; Disease mutation.
FT   SIGNAL       1     20
FT   CHAIN       21    661       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN       24     88       SUSHI 1.
FT   DOMAIN       90    147       SUSHI 2.
FT   DOMAIN      152    209       SUSHI 3.
FT   DOMAIN      212    268       SUSHI 4.
FT   DOMAIN      273    328       SUSHI 5.
FT   DOMAIN      335    390       SUSHI 6.
FT   DOMAIN      395    451       SUSHI 7.
FT   DOMAIN      453    516       SUSHI 8.
FT   DOMAIN      523    579       SUSHI 9.
FT   DOMAIN      581    647       SUSHI 10.
FT   DISULFID     25     76       BY SIMILARITY.
FT   DISULFID     59     87       BY SIMILARITY.
FT   DISULFID     91    135       BY SIMILARITY.
FT   DISULFID    118    146       BY SIMILARITY.
FT   DISULFID    153    197       BY SIMILARITY.
FT   DISULFID    180    208       BY SIMILARITY.
FT   DISULFID    213    255       BY SIMILARITY.
FT   DISULFID    241    267       BY SIMILARITY.
FT   DISULFID    274    316       BY SIMILARITY.
FT   DISULFID    302    327       BY SIMILARITY.
FT   DISULFID    336    378       BY SIMILARITY.
FT   DISULFID    364    389       BY SIMILARITY.
FT   DISULFID    396    439       BY SIMILARITY.
FT   DISULFID    425    450       BY SIMILARITY.
FT   DISULFID    454    505       BY SIMILARITY.
FT   DISULFID    486    515       BY SIMILARITY.
FT   DISULFID    524    567       BY SIMILARITY.
FT   DISULFID    553    578       BY SIMILARITY.
FT   DISULFID    582    636       BY SIMILARITY.
FT   DISULFID    616    646       BY SIMILARITY.
FT   CARBOHYD    162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        617    619       CELL ATTACHMENT SITE.
FT   VARIANT     450    450       C -> F  IN F13B DEFICIENCY).
FT                                /FTId=VAR_007475.
SQ   SEQUENCE   661 AA;  75491 MW;  57A2F8465608572F2 CRC64;

  Query Match             22.0%;  Score  95;  DB 1;  Length 661;
  Best Local Similarity   31.0%;  Pred. No. 3.5e-09;
  Matches   39;  Conservative   19;  Mismatches   58;  Indels   10;  Gaps    5;

Qy       4 CGSPPPILNGRISYYSTPIAVGTVIRYSCSGFFRLIGEKSLICITKDKVDGTWDKPAPKC 63
           |   | :| :| :||    | ::   |   | : |   : |      || | :| |||
Db     274 CPPPPIPINSKIQTHSTTYRHGEIVHIECE.NPEFHGSAEIRC----EDGKWTEP PKC 327

Qy      64 EYFNKYSSCPE-PIVPGGYKIRGSTPYPHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 122
           :  :| | | :  |      |  | :|| ||:|||: : :::|  : | |     |
Db     328 IEGQEKVACEEPPFIENGAANLHSKIYYNG.KVTYACKSGYILHGSNETTCN EGKW--I 384

Qy     123 RLPTCV 128
           | ||
Db     385 LPPECV 390


RESULT 13
DAF2_MOUSE
ID   DAF2_MOUSE      STANDARD;      PRT;   407 AA.
AC   Q61476;
```

```
ID   01 NOV 1997 (Rel. 35, Created)
DT   01 NOV 1997 (Rel. 35, Last sequence update)
DT   01 NOV 1997 (Rel. 35, Last annotation update)
DE   Complement decay accelerating factor, transmembrane precursor
DE   (DAF TM).
GN   DAF2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=95403982; PubMed=7545711;
RA   Spicer A.P., Seldin M.F., Gendler S.J.;
RT   "Molecular cloning and chromosomal localization of the mouse decay-
RT   accelerating factor genes. Duplicated genes encode
RT   glycosylphosphatidylinositol anchored and transmembrane forms.";
RL   J. Immunol. 155:3079-3091(1995).
CC   -!- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: TESTES, SPLEEN AND LYMPH NODE.
CC   -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION, SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE IN SCR3 (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; L41465; AAB00091.1;
DR   HSSP; P08603; 1HCC.
DR   MGD; MGI:104849; Daf2.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Repeat; Signal; Sushi;
KW   Transmembrane.
FT   SIGNAL       1     39       POTENTIAL.
FT   CHAIN       40    407       COMPLEMENT DECAY ACCELERATING FACTOR,
FT                               TRANSMEMBRANE.
FT   DOMAIN      40    408       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    409   429       POTENTIAL.
FT   DOMAIN      430   467       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40    100       SUSHI 1.
FT   DOMAIN      102   164       SUSHI 2.
FT   DOMAIN      167   226       SUSHI 3.
FT   DOMAIN      229   290       SUSHI 4.
FT   DOMAIN      291   368       SER/THR RICH (BY SIMILARITY).
FT   DISULFID     70    99       BY SIMILARITY.
FT   DISULFID    103   150       BY SIMILARITY.
FT   DISULFID    144   163       BY SIMILARITY.
FT   DISULFID    168   209       BY SIMILARITY.
FT   DISULFID    195   225       BY SIMILARITY.
FT   DISULFID    230   272       BY SIMILARITY.
FT   DISULFID    258   289       BY SIMILARITY.
FT   CARBOHYD    192   192       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    267   267       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   407 AA;  44369 MW;  DDDB2F2D72CEAD40 CRC64;

  Query Match             21.4%;  Score 161;  DB 1;  Length 407;
  Best Local Similarity   35.1%;  Pred. No. 5.6e-09;
  Matches   44;  Conservative   23;  Mismatches   53;  Indels   16;  Gaps    8;

Qy       3 SCSSPPTFLNGRTSVYSHFTAVSVERAVSSGPERHGKSAV  [illegible]

Db     167 SCPNPKDLDNGHIN-IPTGILFGSEINFSCNPGYRLVGLS  [illegible]

Qy      62 [illegible]

Db     224 [illegible]

Qy     117 -NMWGPTRRLPTCV 128
           | : :  | |:
Db     279 VGQWS-SPPPQCI 290


RESULT 14
MCP_HUMAN
ID   MCP_HUMAN       STANDARD;       PRT:   477 AA.
AC   P15529;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Membrane cofactor protein precursor (CD46 antigen) (trophoblast
DE   leucocyte common antigen) (TLX).
GN   MCP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 45-58.
RX   MEDLINE=88286080; PubMed=3260937;
RA   Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,
RA   Rebentisch M.B., Lemons R.S., Seya T., Atkinson J.P.;
RT   "Molecular cloning and chromosomal localization of human membrane
RT   cofactor protein (MCP). Evidence for inclusion in the multigene
RT   family of complement-regulatory proteins.";
RL   J. Exp. Med. 168:181-194(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=94119658; PubMed=8418811;
RA   Cervoni F., Fenichel P., Akhoundi C., Bsi B.L., Fossi B.;
RT   "Characterization of a cDNA clone coding for human testis membrane
RT   cofactor protein (MCP, CD46).";
RL   Mol. Reprod. Dev. 34:107-113(1993).
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91267562; PubMed=2050389;
RA   Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
RA   McKenzie I.F.;
RT   "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
RT   regulator of complement activation.";
RL   Immunogenetics 33:335-344(1991).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=94014456; PubMed=7691939;
RA   Cui W., Bourcade D., Post T., Greenlund A.C., Atkinson J.P.,
RA   Kumar V.;
RT   "Characterization of the promoter region of the membrane cofactor
RT   protein (CD46) gene of the human complement system and comparison to
RT   a membrane cofactor protein like genetic element.";
RL   J. Immunol. 151:4137-4146(1993).
RN   [5]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92289809; PubMed=1601037;
RA   Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.B.;
RT   "Tissue specific and allelic expression of the complement regulator
RT   CD46 is controlled by alternative splicing.";
RL   Eur. J. Immunol. 22:1513-1518(1992).
CC   -!- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
CC       CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
CC       FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
CC       ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
CC       CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
CC       THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
```

```
CC       ACTIVATION AT THE FETO MATERNAL INTERFACE ON THE
CC       SYNCYTIOTROPHOBLAST LAYER OF PLACENTA.
CC    !- SUBCELLULAR LOCATION: Type I membrane protein.
CC    !  ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
CC       GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
CC       EBV B CELLS AND LEUKEMIC CELLS. THE 66 kDa ALPHA ISOFORM AND THE
CC       56 kDa BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
CC       TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
CC       RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
CC       PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
CC       AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
CC       12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
CC    !- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
CC       ERYTHROCYTES AND SOME BONE MARROW CELLS.
CC    !  PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
CC       (PROBABLE).
CC    !- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC    !- DATABASE: NAME=PROW; NOTE=CD quide CD46 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd46.htm".
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; Y00651; CAA68675.1; -.
DR   EMBL; S51940; AAB24802.1; -.
DR   EMBL; M58050; AAA62833.1; -.
DR   EMBL; A18585; CAA01400.1; -.
DR   EMBL; S65879; AAD13968.1; -.
DR   PIR; S01896; S01896.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120920; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Sushi; Alternative splicing.
FT   SIGNAL       1     34
FT   CHAIN       35    377       MEMBRANE COFACTOR PROTEIN.
FT   DOMAIN      35    328       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    329   351       POTENTIAL.
FT   DOMAIN      352   377       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       35    95       SUSHI 1.
FT   DOMAIN       98   158       SUSHI 2.
FT   DOMAIN      161   224       SUSHI 3.
FT   DOMAIN      227   284       SUSHI 4.
FT   DOMAIN      287   311       SER/THR-RICH.
FT   DISULFID     35    80       BY SIMILARITY.
FT   DISULFID     64    94       BY SIMILARITY.
FT   DISULFID     99   141       BY SIMILARITY.
FT   DISULFID    127   157       BY SIMILARITY.
FT   DISULFID    162   210       BY SIMILARITY.
FT   DISULFID    191   224       BY SIMILARITY.
FT   DISULFID    228   270       BY SIMILARITY.
FT   DISULFID    256   284       BY SIMILARITY.
FT   CARBOHYD     83    83       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    114   114       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    273   273       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    286   300       MISSING (IN A SECOND FORM).
FT   VARSPLIC    340   352       MISSING (IN ISOFORM M).
FT   VARSPLIC    353   361       YLQRRKKKG -> DIFKGGRRKGKQMVELNMPLTRLNQPLQ
FT                               QSREAE (IN ISOFORM M).
FT   VARSPLIC    362   377       IYLIDETHREVKFTSL -> KADGCAEYATYQTKSTTPAEQ
FT                               RG (IN ISOFORM B/D/F/H/J/L).
FT   VARSPLIC    362   377       MISSING (IN ISOFORM M AND ISOFORM N).
FT   VARSPLIC    340   361       MISSING (IN ISOFORM N).
FT   VARSPLIC    339   339       I -> IGKQMVELNMPLTRLNQPLQQSREAE (IN
FT                               ISOFORM N).
SQ   SEQUENCE   377 AA;  42247 MW;  2CA6FC.752570B57 CRC64;

  Query Match             21.2%;  Score 194;  DB 1;  Length 377;
  Best Local Similarity   27.9%;  Pred. No. 7.7e-09;
  Matches   38;  Conservative   25;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYSIPT-AVGTVIYYSCS-- --GFRLIGEKSLLITTKDKVDGT 55
           : |  || | ||: ::    :      : ||         | |||| :: |     :
Db     160 VLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC -- GDNSV 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-SIPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           | :  |:|    |  | |:| | :| |    : :  :| | |  | ::|: :: |
Db     216 WEPAVP----KVVKCRFPVVENGKQI-SGFGKEFYFRATVMEGDNSFYAVSTIDV- 271

Qy     115 ANNMWGPTRLPTCVSV 130
           :|: | |  :| |: |
Db     272 SNSTWDPP-VPKCLKV 286


RESULT 15
DAF_PONPY
ID   DAF_PONPY       STANDARD;       PRT:   340 AA.
AC   P49457;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Complement decay-accelerating factor (CD55) (Fragment).
GN   DAF OR CD55.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94110622; PubMed=7506731;
RA   Nickells M.W., Alvarez J.I., Lublin D.M., Atkinson J.P.;
RT   "Characterization of DAF-2, a high-molecular weight form of decay
RT   accelerating factor (DAF; CD55), as a covalently cross-linked dimer
RT   of DAF-1.";
RL   J. Immunol. 152:676-685(1994).
CC   -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC       CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC       C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC       INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
CC       INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC       FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC       THE FORMATION OF C4B2A AND C3BB3. THE AMPLIFICATION CONVERTASES OF
CC       THE COMPLEMENT CASCADE (BY SIMILARITY).
CC    !  SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC       HOMODIMER (MINOR FORM).
CC    !  SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor.
CC    -!- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF 2, SHOWN HERE, AND
CC       DAF-1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC    -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC    -!- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
CC    -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC    -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ---------- ----------
DR   EMBL; S67775; AAC60609.1; -.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
```

```
KW   SMART; SM00032; CCP; 4.
KW   Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW   Alternative splicing; GPI anchor; Sushi.
FT   NON_TER          1      1
FT   CHAIN           -1     347       COMPLEMENT DECAY ACCELERATING FACTOR.
FT   PROPEP         348     460       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN          -1      54       SUSHI 1.
FT   DOMAIN          56     118       SUSHI 2.
FT   DOMAIN         121     180       SUSHI 3.
FT   DOMAIN         183     243       SUSHI 4.
FT   DOMAIN         246     315       SER/THR RICH.
FT   DISULFID        24      54       BY SIMILARITY.
FT   DISULFID        52     104       BY SIMILARITY.
FT   DISULFID        88     117       BY SIMILARITY.
FT   DISULFID       122     183       BY SIMILARITY.
FT   DISULFID       149     179       BY SIMILARITY.
FT   DISULFID       184     226       BY SIMILARITY.
FT   DISULFID       212     242       BY SIMILARITY.
FT   CARBOHYD        54      54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       107     107       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID          312     312       GPI-ANCHOR (BY SIMILARITY).
SQ   SEQUENCE   460 AA;  47180 MW;  D4DB65C058204290 CRC64;

  Query Match             21.2%;  Score 158.5;  DB 1;  Length 460;
  Best Local Similarity   42.3%;  Pred. No. 7.8e-09;
  Matches   44;  Conservative   20;  Mismatches   51;  Indels   19;  Gaps    8;

Qy       8 SCGSPPPFLNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLITCKDKVDGT  WDKP 59
           |:  :|  | ||:|   |  :|  :  | :||:  ::|:|   | ||:     : |:    |  |
Db     121 SCENPGEIKNGQIDVSNGILFGATISFSCNTGYKLFGSTSSFCL  ISGSSVQWSDP 175

Qy      60 APKCEEIKCPEPVIPEGGYKIRGSIP YRHGDSVTFACKINESMNGNKSVWCQANN  117
             |:|    :  || |     |:|    | :  |:|:||   :::| |  |:::  |:
Db     176 LPEC    RETYCPAPVQTTNGIQGKKDHYGYRQSITYACNKGYTMIGEHSIYCIVNDD 231

Qy     118   MW GPTRLPTC 127
               :||   | |
Db     232 EGEWSGPP  PEC 242


RESULT 16
DAF_HUMAN
ID   DAF_HUMAN       STANDARD;       PRT;   481 AA.
AC   P08174; P09679; P78646;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement decay-accelerating factor precursor (CD55 antigen).
GN   DAF OR CR OR CD55.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=87115845; PubMed=2433596;
RA   Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
RA   Nussenzweig V.;
RT   "Cloning of decay-accelerating factor suggests novel use of splicing
RT   to generate two proteins.";
RL   Nature 325:545-549(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervix;
RA   Strausberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 6-481 FROM N.A. (ISOFORM 2).
RX   MEDLINE=87175602; PubMed=2436222;
RA   Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
RA   Leykam J.F., Atkinson J.P., Tykocinski M.L.;
RT   "Cloning and characterization of cDNAs encoding the complete sequence
RT   of decay-accelerating factor of human complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
RN   [4]
RP   SEQUENCE OF 35-481 FROM N.A. (ISOFORM 2).
RC   TISSUE=Hippocampus;
RA   Kumar V.B., Hyung C., Nakra R., Walters M., Susser I., Bernardo A.;
RT   "Decay acceleration factor (DAF; CD 55) in the brain of Alzheimer's
RT   disease patients.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-100 FROM N.A.
RX   MEDLINE=91221256; PubMed=1711208;
RA   Ewulonu U.K., Ravi L., Medof M.E.;
RT   "Characterization of the decay-accelerating factor gene promoter
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
RN   [6]
RP   SEQUENCE OF 35-46.
RC   TISSUE=Urine;
RX   MEDLINE=91291869; PubMed=1712233;
RA   Nakano Y., Sugita Y., Ishikawa Y., Choi N. H., Tobe T., Tomita M.;
RT   "Isolation of two forms of decay accelerating factor (DAF) from human
RT   urine.";
RL   Biochim. Biophys. Acta 1074:326-330(1991).
RN   [7]
RP   GPI-ANCHOR.
RX   MEDLINE=91093238; PubMed=1824699;
RA   Moran P., Raab H., Kohr W.J., Caras I.W.;
RT   "Glycophospholipid membrane anchor attachment. Molecular analysis of
RT   the cleavage/attachment site.";
RL   J. Biol. Chem. 266:1250-1257(1991).
RN   [8]
RP   DISULFIDE BONDS IN SUSHI DOMAINS.
RX   MEDLINE=92305034; PubMed=1377029;
RA   Nakano Y., Sumida K., Kikuta N., Miura N. H., Tobe T., Tomita M.;
RT   "Complete determination of disulfide bonds localized within the short
RT   consensus repeat units of decay-accelerating factor (CD55 antigen).";
RL   Biochim. Biophys. Acta 1116:235-240(1992).
RN   [9]
RP   FUNCTION AS A ECHOVIRUS RECEPTOR.
RX   MEDLINE=95045399; PubMed=7525274;
RA   Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
RA   Almond J.W.;
RT   "Decay-accelerating factor CD55 is identified as the receptor for
RT   echovirus 7 using CELICS, a rapid immuno focal cloning method.";
RL   EMBO J. 13:5070-5074(1994).
RN   [10]
RP   VARIANT BLOOD GROUP DR(A-).
RX   MEDLINE=94425573; PubMed=7519480;
RA   Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
RA   Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
RT   "Molecular basis of reduced or absent expression of
RT   decay-accelerating factor in Cromer blood group phenotypes.";
RL   Blood 84:1276-1282(1994).
CC   -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC       CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC       C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION
CC       INTERACTION OF DAF WITH CELL ASSOCIATED C4B AND C3B POLYPEPTIDES
CC       INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC       FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC       THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
CC       THE COMPLEMENT CASCADE.
CC   -!- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
CC       VIRUSES (ECHOVIRUSES 13, 21, 29 AND 33).
CC   -!- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC       HOMODIMER (MINOR FORM).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/DAF-1 AND 2/DAF-2 (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
CC       TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT
CC       PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
CC       LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
CC       ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
```

```
CC     !  DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC        FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC        ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC     !  PTM: THE SER/THR RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
CC     !  POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
CC        SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
CC        TC(A), DR(A), ES(A), WES(B), UMC AND IFC) AND LOW-INCIDENCE
CC        (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
CC        CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
CC        EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
CC        PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
CC        FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
CC        BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
CC        EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
CC        PHENOTYPE.
CC     !  SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC     !  SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC        (RCA) FAMILY.
CC     !  DATABASE: NAME-PROW; NOTE=CD antide CD55 entry;
CC        WWW-"http://www.ncbi.nlm.nih.gov/prow/cd/cd55.htm".
CC     -!- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ----------
DR   EMBL; M31516; AAA52169.1; -.
DR   EMBL; M30142; AAA52168.1; -.
DR   EMBL; BC001288; AAH01288.1; -.
DR   EMBL; M15799; AAA52167.1; -.
DR   EMBL; U88576; AAB48622.1; -.
DR   EMBL; M64653; AAA52170.1; -.
DR   EMBL; M64656; AAA52170.1; JOINED.
DR   EMBL; S72858; AAC60633.1; -.
DR   PIR; B26359; B26359.
DR   PIR; A26359; A26359.
DR   PIR; S16187; S16187.
DR   PIR; A39101; A39101.
DR   PIR; S23148; S23148.
DR   HSSP; P08604; 1HCC.
DR   MIM; 125240; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW   Alternative splicing; GPI anchor; Signal; Sushi; Polymorphism;
KW   Blood group antigen.
FT   SIGNAL        1     34
FT   CHAIN        35    353       COMPLEMENT DECAY ACCELERATING FACTOR.
FT   PROPEP      354    381       REMOVED IN MATURE FORM.
FT   DOMAIN       35     95       SUSHI 1.
FT   DOMAIN       97    159       SUSHI 2.
FT   DOMAIN      162    221       SUSHI 3.
FT   DOMAIN      224    284       SUSHI 4.
FT   DOMAIN      287    356       SER/THR-RICH.
FT   DISULFID     36     81
FT   DISULFID     65     94
FT   DISULFID     98    145
FT   DISULFID    129    158
FT   DISULFID    163    204
FT   DISULFID    190    220
FT   DISULFID    225    267
FT   DISULFID    253    284
FT   CARBOHYD     95     95       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       353    353       GPI-ANCHOR.
FT   VARSPLIC    362    381       HICFTLFGLLGTLVTMGLLT -> SRPVTQAGMRWCDRSSI.
FT                                QSRTPGFKRSFHFSLPSSWYYRAHVFHVDRFAWDASNHCLA
FT                                DLAKEELRRKYTQVYRLFLVS (IN ISOFORM 1).
FT   VARIANT      52     52       R -> L (IN TC(B) ANTIGEN).
FT                                /FTId-VAR_001997.
FT   VARIANT      52     52       R -> P (IN TC(C) ANTIGEN).
FT                                /FTId-VAR_001998.
FT   VARIANT      82     82       L -> R (IN WES(A) ANTIGEN).
FT                                /FTId-VAR_001999.
FT   VARIANT     199    199       S -> L (IN DR(A-) ANTIGEN)
FT                                /FTId-VAR_002000.
FT   VARIANT     227    227       A -> P (IN CR(A-) ANTIGEN).
FT                                /FTId-VAR_002001.
FT   CONFLICT     80     80       T -> I   (IN REF. 1, 2 AND 4).
FT   CONFLICT     85     85       S -> M   (IN REF. 3).
FT   CONFLICT    187    187       S -> T   (IN REF. 4).
FT   CONFLICT    297    297       Q -> H   (IN REF. 4).
SQ   SEQUENCE   381 AA;  41388 MW;  29138B3EB6B43565E CRC64;

  Query Match             21.1%;  Score 133.5;  DB 1;  Length 381;
  Best Local Similarity   33.3%;  Pred. No. 8.8e-09;
  Matches   45;  Conservative   17;  Mismatches   50;  Indels   23;  Gaps    9;

Qy        3 SCGSPPEILNGRISYYSTP--IAVGTVIEYSCSGTFRLIGEKSLLCITKDKVDGT-- WD 57
            || :|  | ||:|     |  |  |  | :||: :::| |  |  |:    : |:   |
Db      162 SCPNPGEIRNGQI---DVPGGILFGATISFSCNTGYKLFGSTSSFCL---ISGSSVQWS 214

Qy       58 KPAPKCEYPNKYSSCPEPIVPGGYKIRGSIVYRHGDSVIFACKTNFSMNGNKSVWCQAN 116
             | |:|    :  || |       |:|   | :  |||:||  |:|    |::  |
Db      215 DPLPEC---REIYCPAPPQIDNGIIQGERDHYGYRQSVTYACNKGFTMIGEHSIYCIVN 270

Qy      117 N---MW-GPTRLPTC 127
            |   | ||  | |
Db      271 NDECEWSGPP--PEC 283


RESULT 17
CCPH_HSVSA
ID   CCPH_HSVSA     STANDARD;      PRT;   360 AA.
AC   Q01016;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Complement control protein homolog precursor (CCPH).
GN   4 OR CCPH.
OS   Herpesvirus saimiri (strain 11).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus
OX   NCBI_TaxID=10383;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92233688; PubMed=1321287
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
RN   [2]
RP   SIMILARITY TO CCP.
RX   MEDLINE=92260674; PubMed=1316492;
RA   Albrecht J.-C., Fleckenstein B.;
RT   "New member of the multigene family of complement control proteins in
RT   herpesvirus saimiri.";
RL   J. Virol. 66:3937-3940(1992).
CC   -!- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM
CC       ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA).
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ----------
```

```
DR   EMBL; X64446; CAA45626.1;  -.
DR   EMBL; X64446; CAA45627.1;  -.
DR   EMBL; X60283; CAA42823.1;  -.
DR   EMBL; X60283; CAA42822.1;  -.
DR   PIR; B42544; WMBE2E.
DR   PIR; A42544; WMBEIE.
DR   PIR; S24567; S24567.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Signal; Repeat; Sushi; Transmembrane; Alternative splicing;
KW   Glycoprotein.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    360       COMPLEMENT CONTROL PROTEIN HOMOLOG.
FT   DOMAIN       83    143       SUSHI 1.
FT   DOMAIN      146    206       SUSHI 2.
FT   DOMAIN      209    265       SUSHI 3.
FT   TRANSMEM    328    350       POTENTIAL.
FT   DISULFID     84    125       BY SIMILARITY.
FT   DISULFID    111    142       BY SIMILARITY.
FT   DISULFID    147    191       BY SIMILARITY.
FT   DISULFID    175    205       BY SIMILARITY.
FT   DISULFID    210    252       BY SIMILARITY.
FT   DISULFID    238    294       BY SIMILARITY.
FT   CARBOHYD     46     46       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     46     46       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     72     72       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    156    156       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    289    302       RICNGNCTTSMPTQ   - AECACPGSNYPISS (IN
FT                                SHORT ISOFORM).
FT   VARSPLIC    303    360       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   360 AA;  43006 MW;  6278A6C2ECD49669 CRC64;

  Query Match             20.0%;  Score 150;  DB 1;  Length 360;
  Best Local Similarity   29.0%;  Pred. No. 6e-08;
  Matches   40;  Conservative   20;  Mismatches   56;  Indels   22;  Gaps    7;

Qy       4 CGSPPPILNGRIS     YYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITK- DKVDGTW 56
           | :|   :|||| :      ||       |:|| |:|:  : |||  :  |: |   :|| |
Db      84 CSTPQDLLNGRYIVIGNLAY      GSVITYICNSGYSLIGSITSACLLRGGRVD  W 135

Qy      57 DKPAPKCEYENKYSSC PEPIVPGGYKIRGSTPYRHGDSVTFAC  KTNFSMNGNKSVWC 113
             | |:        | | : :  |          | : ||||:|  :|  :: :  |  |
Db     136 IPRCPTCD     IKKCKPPPQIANGTHTNVKDFYYYLDEVTYSCNDGTKLTLIGPSSKIC 191

Qy     114 QANNMWGPTRLPTCVSVF 131
               | |     |  :|
Db     192 SETGSWVPNCETKCPETF 209


RESULT 10
IL3B_MOUSE
ID   IL3B_MOUSE      STANDARD;      PRT:   668 AA.
AC   Q07008;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein
DE   glutamine gamma glutamyltransferase B chain) (Transglutaminase B
DE   chain).
GN   F13B OR CF13B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c/OSN; TISSUE=Liver;
RX   MEDLINE=94224141; PubMed=8499049;
RA   Nonaka M., Matsuda Y., Shiroishi T., Moriwaki K., Nonaka M.,
RA   Natsume-Sakai S.;
RT   "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT   and assignment of the gene to chromosome 1: close evolutionary
RT   relationship to complement factor B.";
RL   Genomics 15:535-542(1993).
CC   -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC   -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC   -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY
CC   -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; D10071; BAA00963.1;  -.
DR   PIR; A46013; A46013.
DR   HSSP; P08603; 1HFI.
DR   MGD; MGI:88379; F13b.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi.
FT   SIGNAL        1     20       BY SIMILARITY.
FT   CHAIN        21    668       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN       24     88       SUSHI 1.
FT   DOMAIN       90    147       SUSHI 2.
FT   DOMAIN      152    209       SUSHI 3.
FT   DOMAIN      212    268       SUSHI 4.
FT   DOMAIN      273    328       SUSHI 5.
FT   DOMAIN      335    390       SUSHI 6.
FT   DOMAIN      395    451       SUSHI 7.
FT   DOMAIN      453    516       SUSHI 8.
FT   DOMAIN      523    579       SUSHI 9.
FT   DOMAIN      581    647       SUSHI 10.
FT   DISULFID     25     76       BY SIMILARITY.
FT   DISULFID     59     87       BY SIMILARITY.
FT   DISULFID     91    145       BY SIMILARITY.
FT   DISULFID    118    146       BY SIMILARITY.
FT   DISULFID    153    197       BY SIMILARITY.
FT   DISULFID    180    208       BY SIMILARITY.
FT   DISULFID    213    255       BY SIMILARITY.
FT   DISULFID    241    267       BY SIMILARITY.
FT   DISULFID    274    316       BY SIMILARITY.
FT   DISULFID    302    327       BY SIMILARITY.
FT   DISULFID    336    378       BY SIMILARITY.
FT   DISULFID    364    389       BY SIMILARITY.
FT   DISULFID    396    439       BY SIMILARITY.
FT   DISULFID    425    450       BY SIMILARITY.
FT   DISULFID    454    505       BY SIMILARITY.
FT   DISULFID    486    515       BY SIMILARITY.
FT   DISULFID    524    567       BY SIMILARITY.
FT   DISULFID    553    578       BY SIMILARITY.
FT   DISULFID    582    636       BY SIMILARITY.
FT   DISULFID    616    646       BY SIMILARITY.
FT   CARBOHYD    162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   668 AA;  76078 MW;  80BC9E00A9E54EA6 CRC64;

  Query Match             19.8%;  Score 149;  DB 1;  Length 668;
  Best Local Similarity   31.7%;  Pred. No. 1.5e-07;
  Matches   40;  Conservative   16;  Mismatches   60;  Indels   10;  Gaps    5;

Qy       4 CGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLC IKEKVDGTWDKPAPKC 63
           | ||  || :|  :||    |  :  |   |  : | :|||       :| | :| |||
Db     274 CPEEPVPLNSKIQPHSTTYHPGPDVHHYCPINFVIKSEELLC     ENGKWEPEPKC 327
```

```
Qy      64 EYFNKYSSCPE PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 123
                :   | : | |   |       |   |  || || |   :|: |: : :|      |  |
Db     328 IEEKEKVACEQPPSVENGVAHFHSEIYYSGDKVTYRCGGGYSLRGSSTITCN-RGRW--T 384

Qy     124 RLPTCV 128
           | ||
Db     385 LPPECV 390


RESULT 19
LEM3_MOUSE
ID   LEM3_MOUSE     STANDARD;      PRT;   768 AA.
AC   Q01102;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP OR GRMP.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340571; PubMed=1378846;
RA   Weller A., Isenmann S., Vestweber D.;
RT   "Cloning of the mouse endothelial selectins. Expression of both E-
RT   and P-selectin is inducible by tumor necrosis factor alpha.";
RL   J. Biol. Chem. 267:15176-15183(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92345617; PubMed=1379089;
RA   Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT   "Molecular cloning and analysis of in vivo expression of murine P-
RT   selectin.";
RL   Blood 80:795-800(1992).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC       AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC       ACTIVATION BY AGONISTS, P SELECTIN IS TRANSPORTED RAPIDLY TO
CC       THE CELL SURFACE.
CC   -!- INDUCTION: BY TNF-ALPHA.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS; MOUSE P-LECTIN LACKS
CC       THE HUMAN SUSHI-2 EQUIVALENT.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M87861; AAA40008.1; -.
DR   EMBL; M72342; AAA37712.1; -.
DR   PIR; A42755; A42755.
DR   HSSP; P16109; 1PSB.
DR   MGD; MGI:98280; Selp.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
FT   SIGNAL       1     41       POTENTIAL.
FT   CHAIN       42    768       P SELECTIN.
FT   DOMAIN      42    709       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   710    733       POTENTIAL.
FT   DOMAIN     734    768       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN     159    195       EGF-LIKE.
FT   DOMAIN     199    258       SUSHI 1.
FT   DOMAIN     261    320       SUSHI 2.
FT   DOMAIN     323    382       SUSHI 3.
FT   DOMAIN     385    444       SUSHI 4.
FT   DOMAIN     447    506       SUSHI 5.
FT   DOMAIN     509    568       SUSHI 6.
FT   DOMAIN     579    638       SUSHI 7.
FT   DOMAIN     641    700       SUSHI 8.
FT   DISULFID    60    158       BY SIMILARITY.
FT   DISULFID   141    150       BY SIMILARITY.
FT   DISULFID   163    174       BY SIMILARITY.
FT   DISULFID   168    183       BY SIMILARITY.
FT   DISULFID   185    194       BY SIMILARITY.
FT   DISULFID   200    244       BY SIMILARITY.
FT   DISULFID   240    257       BY SIMILARITY.
FT   DISULFID   252    306       BY SIMILARITY.
FT   DISULFID   292    319       BY SIMILARITY.
FT   DISULFID   324    368       BY SIMILARITY.
FT   DISULFID   354    381       BY SIMILARITY.
FT   DISULFID   386    430       BY SIMILARITY.
FT   DISULFID   416    443       BY SIMILARITY.
FT   DISULFID   448    492       BY SIMILARITY.
FT   DISULFID   478    505       BY SIMILARITY.
FT   DISULFID   510    554       BY SIMILARITY.
FT   DISULFID   540    567       BY SIMILARITY.
FT   DISULFID   580    624       BY SIMILARITY.
FT   DISULFID   610    637       BY SIMILARITY.
FT   DISULFID   642    686       BY SIMILARITY.
FT   DISULFID   672    699       BY SIMILARITY.
FT   CARBOHYD   398    398       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   603    603       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   654    654       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   661    661       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   679    679       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID      745    745       PALMITATE (BY SIMILARITY)
FT   SITE       756    759       ENDOCYTOSIS SIGNAL (PROBABLE).
FT   CONFLICT   724    724       A -> E (IN REF. 2).
SQ   SEQUENCE   768 AA;  83098 MW;  F517-074D2F66E68 CRC64;

  Query Match             19.7%;  Score 148;  DB 1;  Length 768;
  Best Local Similarity   28.6%;  Pred. No. 2.2e-07;
  Matches   36;  Conservative   21;  Mismatches   49;  Indels   20;  Gaps    6;

Qy      16 SYYSTPIAVGIVIRYSCSGTFRLIGEPSIIC ITKDEVDGTWDKPAPKCEYFNKYSS     71
           |:      |||:  :|||:  | |:|  ::       | | |  | | |:   :
Db     524 SHVHGHFGVGSICHFSCNEDFELIGSENVIC- --TVSGKWSAPPPICKIISLPAPAVR 579

Qy      72 CPEPIVPGGYKIRGSTPYRH-----GDSVT -FACKTNFSMNGNKSVWCQANNMWGPTRL 124
           ||    ||    :|:   :|     | :   | | | |:: | |  |:|:
Db     580 CPAITPPG----QGTMSCQHHLGSFGPNTCYFGCKTGFTLRGANSLRCRASSQWTAV 2 634

Qy     125 PTCVSV 130
           | | :|
Db     635 PMCRAV 640
```

```
Result 20
FHR1_HUMAN
ID   FHR1_HUMAN      STANDARD;      PRT:   330 AA.
AC   Q03591; Q96LI7;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor H related protein 1 precursor (FHR-1) (H factor-like
DE   protein 1) (H factor like 1) (H36).
GN   FHR1 OR HFL1 OR CFHL.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91201892; PubMed=1826708;
RA   Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;
RT   "Cloning of the 1.4 kb mRNA species of human complement factor H
RT   reveals a novel member of the short consensus repeat family related
RT   to the carboxy terminal of the classical 150 kDa molecule.";
RL   J. Immunol. 146:3190-3196(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20245592; PubMed=10781834;
RA   Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;
RT   "Complement factor H: Sequence analysis of 221 kb of human genomic DNA
RT   containing the entire FH, FHR-1 and FHR-3 genes.";
RL   Mol. Immunol. 37:41-52(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hall R.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 4-330 FROM N.A.
RX   MEDLINE=91268081; PubMed=1711847;
RA   Skerka C., Horstmann R.D., Zipfel P.F.;
RT   "Molecular cloning of a human serum protein structurally related to
RT   complement factor H.";
RL   J. Biol. Chem. 266:12015-12020(1991).
RN   [5]
RP   CARBOHYDRATES.
RX   MEDLINE=91123700; PubMed=1825108;
RA   Timmann C., Leippe M., Horstmann R.D.;
RT   "Two major serum components antigenically related to complement factor
RT   H are different glycosylation forms of a single protein with no
RT   factor H-like complement regulatory functions.";
RL   J. Immunol. 146:1265-1270(1991).
RN   [6]
RP   REVIEW.
RX   MEDLINE=94226679; PubMed=8172644;
RA   Zipfel P.F., Skerka C.;
RT   "Complement factor H and related proteins: an expanding family of
RT   complement regulatory proteins?";
RL   Immunol. Today 15:121-126(1994).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN
CC       ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID
CC       METABOLISM.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: LIVER.
CC   -!- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE
CC       SIDE CHAIN AND THE OTHER WITH TWO.
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: STRONG TO FACTOR H.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   -----------------------------------------------------------------------
DR   EMBL; M65292; AAA35946.1; -.
DR   EMBL; M65293; AAA35947.1; -.
DR   EMBL; AL049741; CAB53063.1; -.
DR   EMBL; X56209; CAA39666.1; -.
DR   PIR; A40455; A40455.
DR   PIR; S14604; S14604.
DR   HSSP; P10998; 1VVD.
DR   MIM; 134371; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
KW   Repeat; Glycoprotein; Sushi; Signal; Polymorphism.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19    330       COMPLEMENT FACTOR H RELATED PROTEIN 1.
FT   DOMAIN      22     84       SUSHI 1.
FT   DOMAIN      86    141       SUSHI 2.
FT   DOMAIN     146    202       SUSHI 3.
FT   DOMAIN     207    263       SUSHI 4.
FT   DOMAIN     265    328       SUSHI 5.
FT   CARBOHYD   126    126       N-LINKED (GLCNAC...).
FT   CARBOHYD   194    194       N-LINKED (GLCNAC...).
FT   VARIANT    157    157       Y -> H.
FT                               /FTId=VAR_001980.
FT   VARIANT    159    159       V -> I.
FT                               /FTId=VAR_001981.
FT   VARIANT    175    175       Q -> E.
FT                               /FTId=VAR_001982.
FT   CONFLICT    71     71       I -> N (IN REF. 4)
SQ   SEQUENCE   330 AA;  37661 MW;  8DC0D3F92AB5EL45 CRC64;

  Query Match             19.6%;  Score 147;  DB 1;  Length 330;
  Best Local Similarity   27.9%;  Pred. No. 1.1e-07;
  Matches   48;  Conservative   22;  Mismatches   61;  Indels   10;  Gaps    5;

Qy        3 SCGSPPEINGRI-SYYSIPIAVGIYIRYSPSGIEELIGEKSLIGKIGRVIMTWDPAP 61
            |  :|| : |  | |    :   |  :|| |   : : |:  :::|     :| | :| |
Db      146 SCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCL     NGNWTEPPP 199

Qy       62 PIEYPNKYSSPTPEPIVPGYYEIPGSIPYDPGSGVIIAVKENIMNGNEVWPGANNMWG  120
            :|:          | ||  |        : |    || : |:  :  : |:| : |: |  |
Db      200 QCKQSGGGGGGGGGGGNGGGISEPGSVVAGASSVEYGGPNIVGGNSRIIGR NGQWSE 258

Qy      121  PTKLPTCV 128
             |  |  ||
Db      259 PEKCLHPCV 267


RESULT 21
LEM3_RAT
ID   LEM3_RAT        STANDARD;      PRT:   768 AA.
AC   P98106;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=94333817; PubMed=7520013;
RA   Auchampach J.A., Oliver M.G., Anderson D.C., Manning A.M.;
RT   "Cloning, sequence comparison and in vivo expression of the gene
RT   encoding rat P-selectin.";
RL   Gene 145:251-255(1994).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       SIALYLATED OLIGOSACCHARIDES AND MEDIATES THE
```

```
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC    !  SUBCELLULAR LOCATION: Type I membrane protein.
CC    !  TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED: SPLEEN,
CC       LUNG, BRAIN, LIVER, HEART, KIDNEY, THYMUS, SMALL INTESTINE.
CC    !  INDUCTION: BY ACUTE INFLAMATION (PROBABLE).
CC    !  SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC    !  SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC    !  SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC    !  SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS; RAT P-LECTIN LACKS THE
CC       HUMAN SUSHI-2 EQUIVALENT.
CC   -!- ---------- ---------- ---------- --- ----- ----------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -!- ---------- ---------- ---------- --- ----- ----------------------------------
DR   EMBL; L23088; AAA60325.1; -.
DR   HSSP; P16109; 1PSB.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00443; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
FT   SIGNAL        1     41       POTENTIAL.
FT   CHAIN        42    768       P-SELECTIN
FT   DOMAIN       42    709       EXTRACELLULAR (POTENTIAL)
FT   TRANSMEM    710    733       POTENTIAL.
FT   DOMAIN      734    768       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      159    195       EGF-LIKE.
FT   DOMAIN      199    258       SUSHI 1.
FT   DOMAIN      261    320       SUSHI 2.
FT   DOMAIN      323    382       SUSHI 3.
FT   DOMAIN      385    444       SUSHI 4.
FT   DOMAIN      447    506       SUSHI 5.
FT   DOMAIN      509    568       SUSHI 6.
FT   DOMAIN      579    638       SUSHI 7.
FT   DOMAIN      641    700       SUSHI 8.
FT   DISULFID     60    158       BY SIMILARITY.
FT   DISULFID    131    150       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    185    194       BY SIMILARITY.
FT   DISULFID    200    244       BY SIMILARITY.
FT   DISULFID    230    257       BY SIMILARITY.
FT   DISULFID    262    306       BY SIMILARITY.
FT   DISULFID    292    319       BY SIMILARITY.
FT   DISULFID    324    368       BY SIMILARITY.
FT   DISULFID    354    381       BY SIMILARITY.
FT   DISULFID    386    430       BY SIMILARITY.
FT   DISULFID    416    443       BY SIMILARITY.
FT   DISULFID    448    492       BY SIMILARITY.
FT   DISULFID    478    505       BY SIMILARITY.
FT   DISULFID    510    554       BY SIMILARITY.
FT   DISULFID    540    567       BY SIMILARITY.
FT   DISULFID    580    624       BY SIMILARITY.
FT   DISULFID    610    637       BY SIMILARITY.
FT   DISULFID    642    686       BY SIMILARITY.
FT   DISULFID    672    699       BY SIMILARITY.
FT   CARBOHYD     45     45       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    107    107       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    654    654       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    661    661       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       745    745       PALMITATE (BY SIMILARITY).
FT   STOP        756    759       ENDOCYTOSIS SIGNAL (PROBABLE).
SQ   SEQUENCE   768 AA;  83517 MW;  26FD7E8A3F3F1416 CRC64;

  Query Match             19.5%;  Score 145.5;  DB 1;  Length 768;
  Best Local Similarity   28.7%;  Pred. No. 3.1e-07;
  Matches   43;  Conservative   20;  Mismatches   94;  Indels   33;  Gaps    9;

Qy        5 GSPP--------PIL----NGRI--SYYSTPIAVGTVIRYSCSGIFRLIGEKSLICIKDK 51
            |||         | :      | :  |:      :|     :  :  | |:   ::: |
Db      560 GSPPMCLAIKCPEIFAPEQGSLDCSHVHGEASVGSICHFSCNEEFELLGSRNVEC--RT 617

Qy       52 VDGTWDEPAPKCEYFNKYS----SCPEPIVAGVKIRGSVPYRH---GDSVTFACK 100
            | | |  | | |:           ||    ||    :|:   ||     | : |  | ||
Db      618 VSGRWSAPPPTCKGVTSLPVPSVRCPALTPG---QGIMSCRHHLESFGPNTICYFGCK 674

Qy      101 TNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
            | |:: :  |: | |:  |     | | :|
Db      675 TGFTLRGANSLRCGASGQWTAV-TPVCRAV 703


RESULT 22
LFC_CARRO
ID   LFC_CARRO       STANDARD;       PRT;  1019 AA.
AC   Q26422;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Limulus clotting factor C precursor (EC 3.4.21.84) (FC).
OS   Carcinoscorpius rotundicauda (Southeast Asian horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Carcinoscorpius.
OX   NCBI_TaxID=6848;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=95268506; PubMed=7538401;
RA   Ding J.L., Navas M.A. III, Ho B.;
RT   "Molecular cloning and sequence analysis of factor C cDNA from the
RT   Singapore horseshoe crab, Carcinoscorpius rotundicauda.";
RL   Mol. Mar. Biol. Biotechnol. 4:98-103(1995).
CC   -!- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC       SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC       ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC       FORM CATALYZES THE ACTIVATION OF FACTOR B.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg | Ser-104 and
CC       124-Ile-|-Ile-125 bonds in Limulus clotting factor B to form
CC       activated factor B. Cleavage of Pro-Arg-|-Xaa bonds in synthetic
CC       substrates.
CC   -!- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC       LIPOPOLYSACCHARIDES AND CHYMOTRYPSIN (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC       BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -!- ---------- ---------- ---------- --- ----- ----------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; S77063; AAB34361.1; -.
DR   HSSP; P00734; 1UVU.
DR   MEROPS; S01.219; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR004043; LCCL.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50820; LCCL; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Signal; Lectin; Hemolymph clotting;
KW   Glycoprotein; Cell adhesion; EGF-like domain; Sushi; Repeat.
FT   SIGNAL        1     25       BY SIMILARITY.
FT   CHAIN        26   1019       LIMULUS CLOTTING FACTOR C.
FT   CHAIN        26    690       LIMULUS CLOTTING FACTOR C, HEAVY CHAIN.
FT   CHAIN       691   1019       LIMULUS CLOTTING FACTOR C, LIGHT CHAIN.
FT   CHAIN       691    762       LIMULUS CLOTTING FACTOR C, A CHAIN.
FT   CHAIN       763   1019       LIMULUS CLOTTING FACTOR C, B CHAIN.
FT   DOMAIN      102    137       EGF-LIKE.
FT   DOMAIN      142    195       SUSHI 1.
FT   DOMAIN      200    254       SUSHI 2.
FT   DOMAIN      260    321       SUSHI 3.
FT   DOMAIN      325    421       LCCL.
FT   DOMAIN      436    568       C-TYPE LECTIN.
FT   DOMAIN      576    634       SUSHI 4.
FT   DOMAIN      685    748       SUSHI 5.
FT   DOMAIN      763   1019       SERINE PROTEASE.
FT   ACT_SITE    809    809       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    865    865       CHARGE RELAY SYSTEM (BY SIMILARITY)
FT   ACT_SITE    966    966       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   BINDING     960    960       SUBSTRATE (BY SIMILARITY).
FT   DOMAIN      643    689       PRO-RICH.
FT   DISULFID    106    118       BY SIMILARITY.
FT   DISULFID    112    125       BY SIMILARITY.
FT   DISULFID    127    136       BY SIMILARITY.
FT   DISULFID    436    447       BY SIMILARITY.
FT   DISULFID    464    564       BY SIMILARITY.
FT   DISULFID    538    556       BY SIMILARITY.
FT   DISULFID    794    810       BY SIMILARITY.
FT   DISULFID    932    951       BY SIMILARITY.
FT   DISULFID    962    996       BY SIMILARITY.
FT   CARBOHYD    523    523       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    740    740       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    767    767       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    912    912       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1019 AA;  112429 MW;  918A1ED8B817B6C3 CRC64;

  Query Match             19.2%;  Score 144;  DB 1;  Length 1019;
  Best Local Similarity   24.5%;  Pred. No. 7.5e-07;
  Matches   38;  Conservative   31;  Mismatches   50;  Indels   36;  Gaps    8;

Qy          2 ISC-----GSPPPIL---         NGKISYYSIPIAVGIVIRYSCSGIRLIGEKSL 44
              |||      || :              :|:: |  :  |  :|:|:   : |||::|
Db        1?? ...NC_W..........AMV...........KVNA....M....A......................2?9

Qy         45 LCIIKDKVDGIWDKPAPKCEYFNKYSSCP  EPIVPGGYKIR        GSIEYRHGDSV 95
                |  : :| :  |:|:            || :|:    |:::          :  |   |
Db        240 TC----QGNGQWNGQIPQCK  NLVFCPDLDPVNHAEHKVKIGVEQKYGQFP  GSIEV 290

Qy         96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
                |: |  |: : |  :: |  :  |  :: |:||  |
Db        291 TYTCSGNYFLMGFDILKCNPDGSWSGSQ PSCVKV 324

RESULT 23
LFC_TACTR
ID   LFC_TACTR      STANDARD;      PRT;  1019 AA.
AC   P28175;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Limulus clotting factor C precursor (EC 3.4.21.84) (FC).
OS   Tachypleus tridentatus (Japanese horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Tachypleus.
OX   NCBI_TaxID=6853;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91177916; PubMed=2007602;
RA   Muta T., Miyata T., Misumi Y., Tokunaga F., Nakamura T., Toh Y.,
RA   Ikehara Y., Iwanaga S.;
RT   "Limulus factor C. An endotoxin-sensitive serine protease zymogen
RT   with a mosaic structure of complement-like, epidermal growth
RT   factor-like, and lectin-like domains.";
RL   J. Biol. Chem. 266:6554-6561(1991).
CC   -!- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC       SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC       ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC       FORM CATALYZES THE ACTIVATION OF FACTOR B.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg-|-Ser-104 and
CC       124-Ile-|-Ile-125 bonds in Limulus clotting factor B to form
CC       activated factor B. Cleavage of -Pro-Arg-|-Xaa- bonds in synthetic
CC       substrates.
CC   -!- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC       LIPOPOLYSACCHARIDES AND CHYMOTRYPSIN.
CC   -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC       BY A DISULFIDE BOND.
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D90271; BAA14315.1; -.
DR   EMBL; D90272; BAA14316.1; -.
DR   PIR; A38738; A38738.
DR   PIR; B38738; B38738.
DR   HSSP; P00734; 1UVU.
DR   MEROPS; S01.219; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR004043; LCCL.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001304; lectin_c.
```

```
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50820; LCCL; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Signal; Alternative splicing; Lectin;
KW   Hemolymph clotting; Glycoprotein; Cell adhesion; EGF like domain;
KW   Sushi; Repeat.
FT   SIGNAL       1     25
FT   CHAIN       26   1019       LIMULUS CLOTTING FACTOR C.
FT   CHAIN       26    690       LIMULUS CLOTTING FACTOR C, HEAVY CHAIN.
FT   CHAIN      691   1019       LIMULUS CLOTTING FACTOR C, LIGHT CHAIN.
FT   CHAIN      691    762       LIMULUS CLOTTING FACTOR C, A CHAIN.
FT   CHAIN      763   1019       LIMULUS CLOTTING FACTOR C, B CHAIN.
FT   DOMAIN     102    137       EGF LIKE.
FT   DOMAIN     142    195       SUSHI 1.
FT   DOMAIN     200    254       SUSHI 2.
FT   DOMAIN     260    321       SUSHI 3.
FT   DOMAIN     325    421       LCCL.
FT   DOMAIN     436    568       C TYPE LECTIN.
FT   DOMAIN     576    634       SUSHI 4.
FT   DOMAIN     685    748       SUSHI 5.
FT   DOMAIN     763   1019       SERINE PROTEASE.
FT   ACT_SITE   809    809       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   865    865       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   966    966       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   BINDING    960    960       SUBSTRATE (BY SIMILARITY).
FT   DOMAIN     643    689       PRO-RICH.
FT   DISULFID   106    118       BY SIMILARITY.
FT   DISULFID   112    125       BY SIMILARITY.
FT   DISULFID   127    136       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   464    564       BY SIMILARITY.
FT   DISULFID   538    556       BY SIMILARITY.
FT   DISULFID   794    810       BY SIMILARITY.
FT   DISULFID   932    951       BY SIMILARITY.
FT   DISULFID   962    996       BY SIMILARITY.
FT   CARBOHYD   523    523       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   740    740       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   767    767       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   912    912       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   492    498       LTTIWIG -> TDNVTAT (IN SHORT ISOFORM).
FT   VARSPLIC   499   1019       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1019 AA;  112346 MW;  5BC2864C6715289B CRC64;

  Query Match          18.9%;  Score 142;  DB 1;  Length 1019;
  Best Local Similarity   24.5%;  Pred. No. 1.2e-06;
  Matches   88;  Conservative   81;  Mismatches   50;  Indels   36;  Gaps    8;

Qy       2 ISC--  --  GSPPPIL----  --  - NGRISYYSIPIAVGTVIRYSCSGTFRLIGEKSL 44
           |||          ||  :           :|:::  |  :  |  :|:||   : |||:::|
Db     180 ISCLPNGQWSSEPPPKCIREGAKVSSPEHGKVNAPSGNMIEGATLRFSCDSPYYLIGQETL 239

Qy      45 LCIIKDKVDCTWDKPAPKCEYFNKYSSCP--EPIVPGGYKIR-------GSTPYRHGDSV 95
             |    : :| |    |:|:   |   ||  :|:    ::::       |  |   |  |
Db     240 TC    QGNGQWSGQIPQCK---KLVFCPDLDPVNHAEHQVKIGVEQKYGQFP--QGTEV 290

Qy      96 TFACKTNFSMNGNKSVWCQANNMWGPTRLPTCVSV 130
           |: |  |: : |  :: |  :  |  :: |:|| |
Db     291 TYICSGNYFLMGENTLKCNPDGSWSGSQ PSCVKV 324
```

```
RESULT 24
DAF_CAVPO
ID   DAF_CAVPO       STANDARD;       PRT;   [C] AA.
AC   Q60401; Q60402; Q60403; Q60404; Q60405; Q60406; P97254; P97255;
AC   P97256; Q9WT19;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement decay-accelerating factor precursor.
GN   DAF.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=HARTLEY; TISSUE=Spleen;
RX   MEDLINE=95403978; PubMed=7545710;
RA   Nonaka M., Miwa T., Okada N., Nonaka M., Okada H.;
RT   "Multiple isoforms of guinea pig decay-accelerating factor (DAF)
RT   generated by alternative splicing.";
RL   J. Immunol. 155:3037-3048(1995).
CC   -!- FUNCTION: PREVENTS THE FORMATION AND/OR ACCELERATES THE
CC       DISSOCIATION OF C3 CONVERTASE.
CC   -!- SUBCELLULAR LOCATION: CLASS GPI: ATTACHED TO THE MEMBRANE BY A
CC       GPI-ANCHOR
CC   -!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST SIX FORMS OF DAF
CC       TRANSCRIPTS (DISTRIBUTED IN FOUR CLASSES: GPI, TCL, TCS AND SEC)
CC       GENERATED BY ALTERNATIVE SPLICING. ONE GPI-ANCHORED FORM (GDAB
CC       GPI), TWO FORMS WITH LONGER CYTOPLASMIC REGION (GDAB-TCL AND
CC       GDABC-TCL). TWO FORMS WITH SHORTER CYTOPLASMIC REGION (GDA-TCS AND
CC       GDAB-TCS) AND ONE SECRETED FORM (GDAB-SEC).
CC   -!- TISSUE SPECIFICITY: ALL THE ISOFORMS ARE WIDELY EXPRESSED. GPI AND
CC       TCS ARE THE MAJOR FORMS, WHEREAS SEC IS MINOR AND TCL IS ONLY
CC       PRESENT IN TRACE LEVELS.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------------------------------------- --- - ---------- - - -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------- --- - ---------- - - -----
DR   EMBL; D55667; BAA09514.1; -.
DR   EMBL; D55656; BAA09514.1; JOINED.
DR   EMBL; D55657; BAA09514.1; JOINED.
DR   EMBL; D55658; BAA09514.1; JOINED.
DR   EMBL; D55659; BAA09514.1; JOINED.
DR   EMBL; D55660; BAA09514.1; JOINED.
DR   EMBL; D55661; BAA09514.1; JOINED.
DR   EMBL; D55662; BAA09514.1; JOINED.
DR   EMBL; D55663; BAA09514.1; JOINED.
DR   EMBL; D55664; BAA09514.1; JOINED.
DR   EMBL; D55667; BAA09515.1; -.
DR   EMBL; D55656; BAA09515.1; JOINED.
DR   EMBL; D55657; BAA09515.1; JOINED.
DR   EMBL; D55658; BAA09515.1; JOINED.
DR   EMBL; D55659; BAA09515.1; JOINED.
DR   EMBL; D55660; BAA09515.1; JOINED.
DR   EMBL; D55661; BAA09515.1; JOINED.
DR   EMBL; D55662; BAA09515.1; JOINED.
DR   EMBL; D55663; BAA09515.1; JOINED.
DR   EMBL; D55664; BAA09515.1; JOINED.
DR   EMBL; D55665; BAA09515.1; JOINED.
DR   EMBL; D55667; BAA09516.1; -.
DR   EMBL; D55656; BAA09516.1; JOINED.
DR   EMBL; D55657; BAA09516.1; JOINED.
```

```
DR   EMBL; D55658; BAA09516.1; JOINED.
DR   EMBL; D55659; BAA09516.1; JOINED.
DR   EMBL; D55660; BAA09516.1; JOINED.
DR   EMBL; D55661; BAA09516.1; JOINED.
DR   EMBL; D55662; BAA09516.1; JOINED.
DR   EMBL; D55663; BAA09516.1; JOINED.
DR   EMBL; D55664; BAA09516.1; JOINED.
DR   EMBL; D55665; BAA09516.1; JOINED.
DR   EMBL; D55667; BAA09517.1; -.
DR   EMBL; D55655; BAA09517.1; JOINED.
DR   EMBL; D55657; BAA09517.1; JOINED.
DR   EMBL; D55658; BAA09517.1; JOINED.
DR   EMBL; D55659; BAA09517.1; JOINED.
DR   EMBL; D55660; BAA09517.1; JOINED.
DR   EMBL; D55661; BAA09517.1; JOINED.
DR   EMBL; D55662; BAA09517.1; JOINED.
DR   EMBL; D55663; BAA09517.1; JOINED.
DR   EMBL; D55664; BAA09517.1; JOINED.
DR   EMBL; D55665; BAA09517.1; JOINED.
DR   EMBL; D55666; BAA09517.1; JOINED.
DR   EMBL; D55667; BAA09518.1; -.
DR   EMBL; D55656; BAA09518.1; JOINED.
DR   EMBL; D55657; BAA09518.1; JOINED.
DR   EMBL; D55658; BAA09518.1; JOINED.
DR   EMBL; D55659; BAA09518.1; JOINED.
DR   EMBL; D55660; BAA09518.1; JOINED.
DR   EMBL; D55661; BAA09518.1; JOINED.
DR   EMBL; D55662; BAA09518.1; JOINED.
DR   EMBL; D55663; BAA09518.1; JOINED.
DR   EMBL; D55664; BAA09518.1; JOINED.
DR   EMBL; D55665; BAA09518.1; JOINED.
DR   EMBL; D55666; BAA09518.1; JOINED.
DR   EMBL; D55667; BAA09519.1; -.
DR   EMBL; D55656; BAA09519.1; JOINED.
DR   EMBL; D55657; BAA09519.1; JOINED.
DR   EMBL; D55658; BAA09519.1; JOINED.
DR   EMBL; D55659; BAA09519.1; JOINED.
DR   EMBL; D55660; BAA09519.1; JOINED.
DR   EMBL; D55661; BAA09519.1; JOINED.
DR   EMBL; D55662; BAA09519.1; JOINED.
DR   EMBL; D55663; BAA09519.1; JOINED.
DR   EMBL; D55664; BAA09519.1; JOINED.
DR   EMBL; D55666; BAA09519.1; JOINED.
DR   EMBL; D49416; BAA08396.1; -.
DR   EMBL; D49417; BAA08397.1; -.
DR   EMBL; D49418; BAA08398.1; -.
DR   EMBL; D49419; BAA08399.1; -.
DR   EMBL; D49420; BAA08400.1; -.
DR   EMBL; D49421; BAA08401.1; -.
DR   HSSP; P08603; 1HCC.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
KW   Complement pathway; Glycoprotein; Membrane; Repeat; GPI anchor;
KW   Alternative splicing; Signal; Sushi.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    451       COMPLEMENT DECAY ACCELERATING FACTOR.
FT   PROPEP      452    507       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       32     93       SUSHI 1.
FT   DOMAIN       94    158       SUSHI 2.
FT   DOMAIN      159    220       SUSHI 3.
FT   DOMAIN      221    283       SUSHI 4.
FT   DOMAIN      284    458       SER/THR RICH.
FT   LIPID       451    451       GPI ANCHOR (BY SIMILARITY).
FT   CARBOHYD     57     57       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    493    493       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC     82    334       MISSING (IN ISOFORM GDA TCS).
FT   VARSPLIC    335    366       MISSING (IN ISOFORM GDAB GPI, ISOFORM
FT                                GDAB TCL, ISOFORM GDA TCS, ISOFORM GDAB
FT                                TCS AND ISOFORM GDAB SEC).
FT   VARSPLIC    399    424       MISSING (IN ISOFORM GDAB GPI AND GDAB
FT                                SEC).
FT   VARSPLIC    425    458       MISSING (IN ISOFORM GDAB GPI, ISOFORM
FT                                GDAB TCL, ISOFORM GDABC TCL AND ISOFORM
FT                                GDAB SEC).
FT   VARSPLIC    459    507       THVYKVDSFACGASNHWLADIAKEDLRRDESNAQNISSLEQ
FT                                VLGAAQTQ - GHMCIKEIVILLVVLVIIG (IN
FT                                ISOFORM GDAB GPI).
FT   VARSPLIC    459    507       THVYKVDSFACGASNHWLADIAKEDLRRDESNAQNISSLEQ
FT                                VLGAAQTQ - DTCV (IN ISOFORM GDAB SEC).
FT   VARSPLIC    459    507       THVYKVDSFACGASNHWLADIAKEDLRRDESNAQNISSLEQ
FT                                VLGAAQTQ - - ANMRHMRTKNPMLYVIT (IN ISOFORM
FT                                GDA TCS AND ISOFORM GDAB TCS).
SQ   SEQUENCE   507 AA;  55263 MW;  D25BBB7749425210 CRC64;

  Query Match             18.8%;  Score 141.5;  DB 1;  Length 507;
  Best Local Similarity   28.9%;  Pred. No. 6.3e-07;
  Matches   87;  Conservative   18;  Mismatches   62;  Indels   11;  Gaps    6;

Qy       4 CGSPPPIINGRISYYSIPIAVGIYIYYSCSGTFRLIGEKSLIYITKDKVIQHWSKTAPKC 63
           | :|  :::|| ||:  :|| : :|: | :||   :|| || |  |: |    | |    | |
Db     162 CPNPGELVNGNIN-VITDLLLLGSQIFFSCDEGYRLTGEASAFCMIKGNAVG-WSSSLPTC 219

Qy      64 EYFNKYSSCPE-PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANN--MW 119
                 ||| | :  |  :     | :   ||: |   | : |   : |   :    |
Db     220 IKI    ICPEPPQIENGRIVNEDTTYFYRHVVTYECNKGFVLTGESRIICVRGDQVCEW 275

Qy     120 GPTRLPTC 127
              |||
Db     276 SDPP-PTC 282


RESULT 25
C4BB_HUMAN
ID   C4BB_HUMAN     STANDARD;      PRT;   252 AA.
AC   P20851;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein beta chain precursor.
GN   C4BPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93315479; PubMed=8325877;
RA   Hillarp A., Pardo-Manuel F., Ruiz R., de Cordoba S., Dahlback B.;
RT   "The human C4b-binding protein beta-chain gene.";
RL   J. Biol. Chem. 268:15017-15023(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90138962; PubMed=2300577;
RA   Hillarp A., Dahlback B.;
RT   "Cloning of cDNA coding for the beta chain of human complement
RT   component C4b-binding protein: sequence homology with the alpha
RT   chain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1183-1187(1990).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. IT ALSO
CC       INTERACTS WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID P
CC       COMPONENT. THE BETA CHAIN BINDS PROTEIN S.
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC       OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC       BETA CHAIN, A 530 kDa HOMOHEPTAMER OF ALPHA CHAINS OR A 500 kDa
CC       COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC       THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC       BINDING SITE FOR C4B AT THE END.
CC   -!- SIMILARITY: TO C4BP ALPHA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L11244; AAA35615.1; -.
DR   EMBL; L11245; AAA35616.1; -.
DR   EMBL; M29964; AAB59520.1; -.
DR   PIR; A44877; A44877.
DR   PIR; A47107; A47107.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120831; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW   Polymorphism.
FT   SIGNAL        1     17
FT   CHAIN        18    252       C4B BINDING PROTEIN BETA CHAIN.
FT   DOMAIN       22     77       SUSHI 1.
FT   DOMAIN       80    135       SUSHI 2.
FT   DOMAIN      138    192       SUSHI 3.
FT   DISULFID     23     63       BY SIMILARITY.
FT   DISULFID     49     76       BY SIMILARITY.
FT   DISULFID     81    121       BY SIMILARITY.
FT   DISULFID    107    134       BY SIMILARITY.
FT   DISULFID    139    179       BY SIMILARITY.
FT   DISULFID    165    191       BY SIMILARITY.
FT   DISULFID    202    202       INTERCHAIN (WITH ALPHA CHAIN)
FT                                (POTENTIAL).
FT   DISULFID    216    216       INTERCHAIN (WITH ALPHA CHAIN)
FT                                (POTENTIAL).
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC.     ) (POTENTIAL)
FT   CARBOHYD    117    117       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     198    198       P -> S (IN DBSNP:1804226).
FT                                /FTId=VAR_012039.
SQ   SEQUENCE   252 AA;  28457 MW;  0F9CC6406705E2E7 CRC64;

  Query Match           18.8%;  Score 141;  DB 1;  Length 252;
  Best Local Similarity   26.0%;  Pred. No. 3.4e-07;
  Matches   83;  Conservative   21;  Mismatches   59;  Indels   14;  Gaps    6;

Qy       4 CQSPPPILNGRISYYSTPIAVGTVI RYSCSGTFRLIGEKSLLC IKDKVDGTWDKPAPK 62
           |   ||: |   | :     | ::  | |    : |:|:|:| |          ||    :
Db      23 CPELPPVDN--SIFVAKEVEGQILGIYVCIKGYHLVGKKTLFC----NASKEWDNTTTE 75

Qy      63 CEYFNKYSSCPEPIVPGGYKIRGSTIYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 122
           |    :   ||:|::  | :   | |    | :|| |  :: : |:    |  :: | |
Db      76 C    RLGHCPDPVIVNG-EFSSSGDVNVSDKITFMCNDHYILKGSNRSQCLEDHTWAPP 130

Qy     123 RLPTCVS 129
           | | |
Db     131 FPICKS 136


RESULT 26
CFAB_PIG
ID   CFAB_PIG       STANDARD;      PRT;   151 AA.
AC   Q03710;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B (EC 4.4.21.47) (C3/C5 convertase) (Properdin
DE   factor B) (Fragment).
GN   BF.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91372866; PubMed=1680099;
RA   Peelman L.J., van de Weghe A.R., Coppieters W.R., van Zeveren A.J.,
RA   Bouquet Y.H.;
RT   "Cloning and sequencing of the porcine complement factor B.";
RL   Immunogenetics 34:192-195(1991).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha-chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha-chain to yield C5a and C5b. Both
CC       cleavages take place at the C-terminal of an arginine residue.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M59240; AAA31021.1; -.
DR   HSSP; P02743; 1QUB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; PARTIAL.
DR   PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
DR   PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi.
FT   NON_TER       1      1
FT   DOMAIN        1     58       SUSHI 2.
FT   DOMAIN       63    118       SUSHI 4.
FT   DISULFID      4     46       BY SIMILARITY.
FT   DISULFID     32     59       BY SIMILARITY.
FT   DISULFID     66    106       BY SIMILARITY.
FT   DISULFID     92    119       BY SIMILARITY.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     151    151
SQ   SEQUENCE   151 AA;  16765 MW;  BD247E75047E517F CRC64;

  Query Match           18.6%;  Score 140;  DB 1;  Length 151;
  Best Local Similarity   30.6%;  Pred. No. 2.4e-07;
  Matches   41;  Conservative   15;  Mismatches   62;  Indels   16;  Gaps    4;

Qy       2 ISCQSPPPILNG----RISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLC IKDKVDGTWD 57
           | |     ||     |  ||:    :   | : |   : | | :  |    :| | ||
Db       2 IRCPRPHLFENGEYWPRAPYYN----LSDEISFHCYDGYILRGSANRIC -  QVIGRWD 53

Qy      58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTIYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
                :   |  || | :| | :    | ||  |||| : |    :: |::   ||
Db      54 GQTAICDDGAGY--CPNPGIPIGTRKVG-TQYRLESVTYYCTRGLTLRGSQRRICQEGG 110

Qy     118 MWGPTRLPTCVSVF 131
           |  |  ::|   |
Db     111 SWSGIE-PSCQDSF 123
```

```
RESULT 27
C4BB_RAT
ID   C4BB_RAT        STANDARD;      PRT:   258 AA.
AC   Q63515;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   C4B-binding protein beta chain precursor.
GN   C4BPB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY; TISSUE=Liver;
RX   MEDLINE=97166082; PubMed=9014975;
RA   Hillarp A., Wikstrand S., Dahlback B.;
RT   "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
RT   structural and functional relationships among human, bovine, rabbit,
RT   mouse, and rat proteins.";
RL   J. Immunol. 158:1315-1323(1997).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DECRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. IT
CC       INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID
CC       P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: TO C4BP ALPHA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z50652; CAA90492.1;
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL       1     19       BY SIMILARITY.
FT   CHAIN       20    258       C4B-BINDING PROTEIN BETA CHAIN.
FT   DOMAIN      20     74       SUSHI 1 (ATYPICAL; LACK A CYS).
FT   DOMAIN      77    134       SUSHI 2.
FT   DOMAIN     135    189       SUSHI 3.
FT   DISULFID    46     74       BY SIMILARITY.
FT   DISULFID    78    118       BY SIMILARITY.
FT   DISULFID   136    179       BY SIMILARITY.
FT   DISULFID   162    188       BY SIMILARITY.
FT   DISULFID   217    217       INTERCHAIN (WITH ALPHA CHAIN)
FT                               (POTENTIAL).
FT   CARBOHYD    27     27       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    67     67       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    89     89       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    95     95       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   114    114       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   218    218       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   258 AA;  28641 MW;  646F99908AB5DC1C CRC64;

  Query Match             18.6%;  Score 140;  DB 1;  Length 258;
  Best Local Similarity   27.1%;  Pred. No. 4.3e-07;
  Matches   46;  Conservative   18;  Mismatches   55;  Indels   24;  Gaps    6;

Qy        4 SCGSPPPILNGRISYYSTPIAVGTVIR      YSCSGTFRIIGEKSIICITKDKVDWSTW 56
            || ||| | |     : ||          | |   : |:|:|||:     :    |
Db       10 SCSEPPPVNNS      VEVGKEIEFQCTEGYYLCIKGYHLVGKKSLVFDPSKE  - W 66

Qy       57 PKPAPKCEYGNKYSSCGEPIVPGGYKIRGSTPYRHGDSVEFACKENESMNGNKKVWCQAN 116
            :   |:|          |||:|:  | ||  | |     : | |   : : |:    | :
Db       67 NSTLPEC     EIGIPCTDPVLENG KINSSCFVNSSWSMEP NPSFGICVSNWSQELED 121

Qy      117 NMWGPTRLPICVS 129
            : | |  || | |
Db      122 RIWAPPFLPICKS 133


RESULT 28
CO2_HUMAN
ID   CO2_HUMAN       STANDARD;      PRT:   752 AA.
AC   P06681; Q13904;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase)
GN   C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=87127920; PubMed=2949737;
RA   Bentley D.R.;
RT   "Primary structure of human complement component C2. Homology to two
RT   unrelated protein families.";
RL   Biochem. J. 239:339-345(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93315833; PubMed=8326124;
RA   Ishii Y., Zhu Z.B., Macon K.J., Volanakis J.E.;
RT   "Structure of the human C2 gene.";
RL   J. Immunol. 151:170-174(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Dankers C., Baskin D., Fauci J., Loretz C., Ahearn M.E.,
RA   Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT   "Sequence determination of 300 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 587-752 FROM N.A.
RX   MEDLINE=84144868; PubMed=6199794;
RA   Bentley D.R., Porter R.R.;
RT   "Isolation of cDNA clones for human complement component C2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:1212-1215(1984).
RN   [5]
RP   SEQUENCE OF 694-752 FROM N.A.
RX   MEDLINE=87102880; PubMed=3643061;
RA   Wu L., Morley B.J., Campbell R.D.;
RT   "Cell-specific expression of the human complement protein factor B
RT   gene: evidence for the role of two distinct 5' flanking elements.";
RL   Cell 48:331-342(1987).
RN   [6]
RP   SEQUENCE OF 137-171; 454-466 AND 574-717.
RX   MEDLINE=85038851; PubMed=6149575;
RA   Gagnon J.;
RT   "Structure and activation of complement components C2 and factor B.";
RL   Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:401-409(1984).
RN   [7]
RP   SEQUENCE OF 244-269.
RX   MEDLINE=83308518; PubMed=6555044;
RA   Parkes C., Gagnon J., Kerr M.A.;
RT   "The reaction of iodine and thiol blocking reagents with human
RT   complement components C2 and factor B. Purification and N-terminal
RT   amino acid sequence of a peptide from C2a containing a free thiol
RT   group.";
RL   Biochem. J. 214:201-209(1983).
RN   [8]
```

```
RP   VARIANTS C2D PHE-209 AND ARG-464.
RX   MEDLINE=96215049; PubMed=8621452;
RA   Wetsel R.A., Kulics J., Tokki M.T., Kiepiela P., Akama H.,
RA   Johnson C.A., Densen P., Colten H.R.;
RT   "Type II human complement C2 deficiency. Allele-specific amino acid
RT   substitutions (Ser189 --> Phe; Gly444 --> Arg) cause impaired C2
RT   secretion.";
RL   J. Biol. Chem. 271:5824-5841(1996).
RN   [9]
RP   VARIANT C2D TYR-131.
RX   MEDLINE=98344005; PubMed=9670940;
RA   Zhu Z.B., Atkinson T.P., Volanakis J.E.;
RT   "A novel type II complement C2 deficiency allele in an African
RT   American family.";
RL   J. Immunol. 161:578-584(1998).
CC   !- FUNCTION: COMPONENT C2 WHICH IS PART OF THE CLASSICAL PATHWAY OF
CC       THE COMPLEMENT SYSTEM IS CLEAVED BY ACTIVATED FACTOR C1 INTO TWO
CC       FRAGMENTS: C2B AND C2A. C2A, A SERINE PROTEASE, THEN COMBINES WITH
CC       COMPLEMENT FACTOR 4B TO GENERATE THE C3 OR C5 CONVERTASE.
CC   !- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND
CC       C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH
CC       CLEAVAGES TAKE PLACE AT THE C TERMINAL OF AN ARGININE RESIDUE.
CC   !- DISEASE: DEFECTS IN C2 ARE THE CAUSE OF C2 DEFICIENCY (C2D). THIS
CC       IS AN AUTOSOMAL RECESSIVE DISEASE. DEFICIENT INDIVIDUALS HAVE AN
CC       INCREASED INCIDENCE OF SLE AND SLE-LIKE SYNDROMES,
CC       GLOMERULONEPHRITIS, VASCULITIS AND PYOGENIC INFECTIONS. TYPE I C2D
CC       IS CHARACTERIZED BY COMPLETE LOSS OF THE PROTEIN WHILE TYPE II C2D
CC       IS CHARACTERIZED BY A SELECTIVE BLOCK IN C2 SECRETION.
CC   !- MISCELLANEOUS: C2 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
CC       PROTEIN.
CC   !- SIMILARITY: WITH COMPLEMENT FACTOR B.
CC   !- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15082; AAA59624.1; -.
DR   EMBL; L09708; AAB97607.1; -.
DR   EMBL; L09706; AAB97607.1; JOINED.
DR   EMBL; L09707; AAB97607.1; JOINED.
DR   EMBL; AF019414; AAB67975.1; -.
DR   EMBL; X04481; CAA28169.1; -.
DR   PIR; A25290; C2HU.
DR   PIR; A05289; A05289.
DR   HSSP; P00734; 1VR1.
DR   MEROPS; S01.194; -.
DR   MIM; 217000; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 4.
DR   Pfam; PF00089; trypsin; 2.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease.
KW   Signal; Repeat; Sushi; Disease mutation; Polymorphism.
FT   SIGNAL       1     20
FT   CHAIN       21    752       COMPLEMENT C2.
FT   CHAIN       21    243       COMPLEMENT C2B FRAGMENT.
FT   CHAIN      244    752       COMPLEMENT C2A FRAGMENT.
FT   DOMAIN      23     85       SUSHI 1.
FT   DOMAIN      88    145       SUSHI 2.
FT   DOMAIN     150    205       SUSHI 3.
FT   DOMAIN     254    452       VWFA.
FT   DOMAIN     466    752       SERINE PROTEASE.
FT   ACT_SITE   507    507       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   561    561       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    24     64       BY SIMILARITY.
FT   DISULFID    51     84       BY SIMILARITY.
FT   DISULFID    89    131       BY SIMILARITY.
FT   DISULFID   117    144       BY SIMILARITY.
FT   DISULFID   151    191       BY SIMILARITY.
FT   DISULFID   177    204       BY SIMILARITY.
FT   CARBOHYD    29     29       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   112    112       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   290    290       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   333    333       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   467    467       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   621    621       N-LINKED (GLCNAC...).
FT   CARBOHYD   651    651       N-LINKED (GLCNAC...).
FT   VARIANT    131    131       C -> Y (IN C2D; TYPE II).
FT                               /FTId=VAR_008544.
FT   VARIANT    209    209       S -> F (IN C2D; TYPE II).
FT                               /FTId=VAR_008545.
FT   VARIANT    464    464       G -> R (IN C2D; TYPE II).
FT                               /FTId=VAR_008546.
FT   VARIANT    533    533       F -> L (IN DBSNP:1042664).
FT                               /FTId=VAR_011772.
SQ   SEQUENCE   752 AA;  83267 MW;  5A96A13E700CF444 CRC64;
```

Query Match 18.6%; Score 140; DB 1; Length 752;
Best Local Similarity 28.5%; Pred. No. 1.4e-06;
Matches 37; Conservative 16; Mismatches 61; Indels 16; Gaps 6;

```
Qy       2 ISCGSPPPPILNG----RISYYSTPIAVGTIEYSCSETFRLIGEKSLICITKDKVDGIWD 57
           : | :|    ||    |: |    ||  : :|   | | |    |   : :| ||
Db      87 VRCEAPVSPENGIYTPRLGSY----PVGENVSFECEDGFILRGSPVRQC----RPNGMWD 138

Qy      58 KPAFKCEYFNKYSSCPEPIVPGGYKIRGSTFYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
             |:  |   || | :  |  :|      ||| : : | :|  : |:    || |
Db     139 GETAVCD--NGAGHCPNPGISLG-AVRGSPFFGHGDEVRYRCSSNFVLTGSSEREGQGNG 195

Qy     118 MWGPTRLPTC 127
           :|  |  | |
Db     196 VWSGTE-PIC 204
```

```
RESULT 29
APOH_MOUSE
ID   APOH_MOUSE     STANDARD;      PRT:   345 AA.
AC   Q01339;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo H) (B2GPI)
DE   (Beta(2)GPI) (Activated protein C-binding protein) (APC inhibitor).
GN   APOH OR B2GPI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92372000; PubMed=1339387;
RA   Nonaka M., Matsuda Y., Shiroishi T., Moriwaki K., Natsuume-Sakai S.;
RT   "Molecular cloning of mouse beta 2-glycoprotein I and mapping of the
RT   gene to chromosome 11.";
RL   Genomics 13:1082-1087(1992).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBA/J; TISSUE=Liver;
RX   MEDLINE=94242017; PubMed=7514402;
RA   Sellar G.C., Steel D.M., Zafiropoulos A., Seery L.T.,
RA   Whitehead A.S.;
RT   "Characterization, expression and evolution of mouse beta 2
RT   glycoprotein I (apolipoprotein H).";
RL   Biochem. Biophys. Res. Commun. 200:1521-1528(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Liver;
RA   Kristensen T.;
RT   "Structure of the human beta 2 glycoprotein I gene.";
RC   Submitted (FEB-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10056; BAA00945.1; -.
DR   EMBL; S70439; AAB30089.1; -.
DR   PIR; I51356; I51356.
DR   PIR; A43286; A43286.
DR   HSSP; P02749; 1CFZ.
DR   MGD; MGI:88058; Apoh.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19
FT   CHAIN       20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      22     86       SUSHI 1.
FT   DOMAIN      84    138       SUSHI 2.
FT   DOMAIN     141    201       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     264    345       SUSHI-LIKE.
FT   DISULFID    24     66       BY SIMILARITY.
FT   DISULFID    51     79       BY SIMILARITY.
FT   DISULFID    84    124       BY SIMILARITY.
FT   DISULFID   110    137       BY SIMILARITY.
FT   DISULFID   142    188       BY SIMILARITY.
FT   DISULFID   174    200       BY SIMILARITY.
FT   DISULFID   205    248       BY SIMILARITY.
FT   DISULFID   234    260       BY SIMILARITY.
FT   DISULFID   264    315       BY SIMILARITY.
FT   DISULFID   300    325       BY SIMILARITY.
FT   DISULFID   307    345       BY SIMILARITY.
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   183    183       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   194    194       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   252    252       G -> R (IN REF. 2).
SQ   SEQUENCE   345 AA;  38615 MW;  CB4EBA0EBD51C940 CRC64;

  Query Match             18.2%;  Score 137;  DB 1;  Length 345;
  Best Local Similarity    8.0%;  Pred. No. 1.2e-06;
  Matches   40;  Conservative   14;  Mismatches   56;  Indels   20;  Gaps    6;

Qy      12 NGRIEYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPKCEYFNKYSS 71
           ||: | | |       | :::: | | | | |       :|| |    | |     :
Db      92 NGIVRY-TSFEYPKNISFACNPGFFLNGTESSSKCTE--EGKWSPDIPACARI--PT 147

Qy      72 CPEPIVP     GGYKIKGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPTRLPT 126

          .
```

```
           |: | ||       |:       : |:| | | :|:| :| :: :     :  ||||
Db     142 CPEPPVPKFALLKDYRPSAGNNSLYQDIVVEKCLPHRAMIGNDTVMCTLQGNW TKLIE 199

Qy     127 CVSV--PP 134
           |: |  ||
Db     200 CLEVKCPEP 208


RESULT 30
HIG_DROME
ID   HIG_DROME      STANDARD;      PRT;   958 AA.
AC   Q09101;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   [1] MAS ...... (Rel. 41, Last annotation update)
DE   Locomotion-related protein Hikaru genki precursor.
GN   HIG.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Head;
RX   MEDLINE=93213498; PubMed=8461133;
RA   Hoshino M., Matsuzaki F., Nabeshima Y.-I., Hama C.;
RT   "Hikaru genki, a Drosophila gene identified by abnormal locomotion
RT   in Drosophila, encodes a novel type of protein.";
RL   Neuron 10:395-407(1993).
CC   -!- FUNCTION: HAS A ROLE IN THE DEVELOPMENT OF CNS FUNCTIONS INVOLVED
CC       IN LOCOMOTOR ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN PCC NEURONS AND NEUROBLASTS
CC       IN THE PROCEPHALIC NEUROGENIC REGION IN THE CENTRAL NERVOUS
CC       SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: MOST ABUNDANT DURING AND JUST AFTER NEURONAL
CC       DIFFERENTIATION AND DURING CELL SPECIFICATION OR AXOGENESIS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2 TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13884; BAA02984.1; -.
DR   EMBL; D13885; BAA02985.1; -.
DR   EMBL; D13886; BAA02986.1; -.
DR   EMBL; D13887; BAA02987.1; -.
DR   HSSP; P10908; 1VVD.
DR   FlyBase; FBgn0010114; hiq.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00409; IG; 1.
KW   Glycoprotein; Alternative splicing; Immunoglobulin domain; Repeat;
KW   Sushi; Signal.
FT   SIGNAL       1     31       POTENTIAL.
FT   CHAIN       32    958       LOCOMOTION-RELATED PROTEIN HIKARU GENKI.
FT   DOMAIN     630    709       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     713    769       SUSHI 1.
FT   DOMAIN     772    828       SUSHI 2.
FT   DOMAIN     831    891       SUSHI 3.
```

```
FT   DOMAIN       893    952       SUSHI 4.
FT   SITE         318    320       CELL ATTACHMENT SITE.
FT   DISULFID     714    755       BY SIMILARITY.
FT   DISULFID     741    768       BY SIMILARITY.
FT   DISULFID     773    814       BY SIMILARITY.
FT   DISULFID     800    827       BY SIMILARITY.
FT   DISULFID     832    877       BY SIMILARITY.
FT   DISULFID     863    890       BY SIMILARITY.
FT   DISULFID     894    939       BY SIMILARITY.
FT   DISULFID     922    952       BY SIMILARITY.
FT   CARBOHYD     376    376       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     525    525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     605    605       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     752    752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     789    789       N-LINKED (GLCNAC    ) (POTENTIAL)
FT   VARSPLIC     529    553       MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARSPLIC     892    958       MISSING (IN ISOFORM 1 AND ISOFORM 2).
SQ   SEQUENCE   958 AA;  107027 MW;  4161258E85ABC764 CRC64;

  Query Match             18.2%;  Score 136.5;  DB 1;  Length 958;
  Best Local Similarity   31.6%;  Pred. No. 4e-06;
  Matches   43;  Conservative   16;  Mismatches   58;  Indels   19;  Gaps    8;

Qy       2 ISCGSPPPTLNGRI---SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDK 58
           | |  |  ||||   |     : || :: :||:    |:|| |::|      :| |
Db     880 IQCVMPVAPLNGRIGGTSLSQRRLTVGALVTFSCNDGHSLVGESSIICTE----NGQWSH 885

Qy      59 PAPKCEYFNKYSSCPEP-IVPGGYKIRGSTPYFHGDSVTFACKTNF--SMNG-NKSVWCQ 114
             | :     | || |   | |         | || ::  |:  |  |  |  :   ||
Db     886 SPPECK---SQCPYPGDPPNGLIAFLKFNYDAGDYLSVQCRPGFVQSYEGPPERPKCQ 940

Qy     115 ANNMW GPTRLPTCVS 129
           :  | ||  :| | |
Db     941 PDGRWSGP-MPKCKS 954


RESULT 31
FHR2_HUMAN
ID   FHR2_HUMAN      STANDARD;      PRT;   270 AA.
AC   P36980; Q14410;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor H-related protein 2 precursor (FHR-2) (H factor-like
DE   protein 2) (H factor-like 4) (DDESK59).
GN   FHR2 OR HFL4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92251200; PubMed=1538657;
RA   Skerka C., Timmann C., Horstmann R.D., Zipfel P.F.;
RT   "Two additional human serum proteins structurally related to
RT   complement factor H. Evidence for a family of factor H-related
RT   genes.";
RL   J. Immunol. 148:3313-3318(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95402981; PubMed=7672821;
RA   Skerka C., Moulds J.M., Taillon-Miller P., Hourcade D., Zipfel P.F.;
RT   "The human factor H-related gene 2 (FHR2): structure and linkage to
RT   the coagulation factor XIIIb gene.";
RL   Immunogenetics 42:268-274(1995).
RN   [3]
RP   REVIEW.
RX   MEDLINE=94226679; PubMed=8172644;
RA   Zipfel P.F., Skerka C.;
RT   "Complement factor H and related proteins: an expanding family of
RT   complement regulatory proteins?";
RL   Immunol. Today 15:121-126(1994).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN
CC       ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID
CC       METABOLISM.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       TRUNCATED FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE
CC       SIDE CHAIN, THE OTHER IS NOT GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: STRONG, TO FACTOR H.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64877; CAA46096.1; -.
DR   EMBL; X86564; CAA60375.1; -.
DR   EMBL; X86565; CAA60375.1; JOINED.
DR   EMBL; X86566; CAA60375.1; JOINED.
DR   EMBL; X86567; CAA60375.1; JOINED.
DR   PIR; S24564; S24564.
DR   HSSP; P08603; 1HFI.
DR   MIM; 600889; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Repeat; Glycoprotein; Sushi; Signal; Alternative splicing.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    270       COMPLEMENT FACTOR H-RELATED PROTEIN 2.
FT   DOMAIN       19     85       SUSHI 1.
FT   DOMAIN       86    127       SUSHI 2.
FT   DOMAIN      128    205       SUSHI 3.
FT   DOMAIN      206    270       SUSHI 4.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .).
FT   VARSPLIC    144    171       ISAEKCGPPPIDNGDITSFLLSVYAPG - - S (IN
FT                                TRUNCATED ISOFORM).
FT   CONFLICT     85     85       R -> K - N REF. 2).
SQ   SEQUENCE   270 AA;  30651 MW;  E1B2E9C139B2.7A8 CRC64;

  Query Match             17.9%;  Score 134.5;  DB 1;  Length 270;
  Best Local Similarity   28.5%;  Pred. No. 1.6e-06;
  Matches   39;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy       4 CGSPPPTLNGRI-SYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPK 62
           || |,|| || | |:  :  | |: : | |    ::| |    :| |  :| ||
Db     149 CGPPPIDNGDITSFLLSVYAPGSSVEYQCQNLYQLEGNNQITC-----RNGQWSEPPK 202

Qy      63 C---------EYFNKYSSCPEPIVPGGYKIRGSTPY--- RHGDSVTFACKTNFSMNGNKS 110
           |         |   ||:           | : :       | || | | ||: :    : |
Db     203 CGPPPVVSQEIMEKYN----------IKLKWTNQQKLYSRIGDTVEFVCKSGYHPTSHS 252

Qy     111 VWCQANNMWGPTRLPTC 127
                 |  |    |:|
Db     253 FRAMCQN--GKLVYPSC 267


RESULT 32
APOH_RAT
ID   APOH_RAT       STANDARD;      PRT:   297 AA.
AC   P26644;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB 1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apo-lipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI).
GN   APOH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY;
RX   MEDLINE=89366680; PubMed=2771654;
RA   Aoyama Y., Chan Y.L., Wool I.G.;
RT   "The primary structure of rat beta 2 glycoprotein I.";
RL   Nucleic Acids Res. 17:6401-6401(1989).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; X17561; CAA35056.1.
DR   PIR; S05416; NBRT.
DR   HSSP; P02749; 1CLZ.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Heparin binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    297       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      20     82       SUSHI 1 (INCOMPLETE).
FT   DOMAIN      85     90       SUSHI 2.
FT   DOMAIN      93    153       SUSHI 3.
FT   DOMAIN     156    213       SUSHI 4.
FT   DOMAIN     215    297       SUSHI-LIKE.
FT   DISULFID    46     76       BY SIMILARITY.
FT   DISULFID    62     89       BY SIMILARITY.
FT   DISULFID    94    140       BY SIMILARITY.
FT   DISULFID   126    152       BY SIMILARITY.
FT   DISULFID   167    200       BY SIMILARITY.
FT   DISULFID   186    212       BY SIMILARITY.
FT   DISULFID   216    267       BY SIMILARITY.
FT   DISULFID   252    277       BY SIMILARITY.
FT   DISULFID   259    297       BY SIMILARITY.
FT   CARBOHYD    69     69       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   114    114       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   135    135       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   145    145       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   205    205       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   297 AA;  33104 MW;  9112D9DA6C119D59 CRC64;

  Query Match             17.8%;  Score 134;  DB 1;  Length 297;
  Best Local Similarity   30.2%;  Pred. No. 2e-06;
  Matches   39;  Conservative   15;  Mismatches   55;  Indels   20;  Gaps    6;

Qy      12 NGRISYYSTFTIAVQEKIDKVFSKSKRPLGHPDTLYQEVMVDPVQTDNSMYDPNVFKSGP 71
           || : .  |       :  ::  : | | |  |  |       :| |    | |     :
Db      44 NGVVRKT-IFEFYTNIIGFAQNIQYYLNGISSSKCIE-----EGKWSDFLPVCARI---SI 93

Qy      72 YPEPIVP-----GSYKFDGYEKDNKEYVTFAGFTPNGNKESVWVQANNMWVSPISLPT 126
           || | |       ||    : :|:| |  :|:| || :| | |:  | |  |:|||
Db      94 FPFPFTPKFAALKEYKTEVGNSSFYQDTVVFKCLPHFAMFGNDTVTCTAHGNW  TQLPE 151

Qy     127 CVSV   FP 132
           | |   ||
Db     152 CREVKCPFP 160


RESULT 33
CFAB_HUMAN

ID   CFAB_HUMAN      STANDARD;      PRT;   764 AA.
AC   P00751; O15006; Q29944; Q9BIF5; Q98X92;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B precursor (EC 3.4.21.47) (C4/C5 convertase)
DE   (Properdin factor B) (Glycine-rich beta glycoprotein) (GBG) (PBF2).
GN   BF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ALLELES S; FA AND FB).
RX   MEDLINE=91065702; PubMed=2249879;
RA   Davrinche C., Abbal M., Clerc A.;
RT   "Molecular characterization of human complement factor B subtypes.";
RL   Immunogenetics 32:309-312(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=Liver;
RX   MEDLINE=94237735; PubMed=8181962;
RA   Mejia J.E., Jahn I., de la Salle H., Hauptmann G.;
RT   "Human factor B. Complete cDNA sequence of the FB*S allele.";
RL   Hum. Immunol. 39:49-53(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=Liver;
RX   MEDLINE=94041399; PubMed=8225386;
RA   Schwaeble W., Luettig B., Sobolowski I., Estaller C., Weiss E.H.,
RA   Meyer zum Buschenfelde K.-H., Whaley K., Dippold W.;
RT   "Human complement factor B: functional properties of a recombinant
RT   zymogen of the alternative activation pathway convertase.";
RL   Immunobiology 188:221-242(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ALLELE S).
RX   MEDLINE=94067177; PubMed=8247029;
RA   Horiuchi T., Kim S., Matsumoto M., Watanabe T., Fujita S.,
RA   Volanakis J.E.;
RT   "Human complement factor B: cDNA cloning, nucleotide sequencing,
RT   phenotypic conversion by site-directed mutagenesis and expression.";
RL   Mol. Immunol. 30:1587-1592(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Dankers C., Baskin D., Faust J., Loret C., Ahearn M.E.,
RA   Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT   "Sequence determination of 300 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Jaatinen T., Kaneva J., Poutanen K.E., Saarinen Pihkala R.,
RA   Lokki M.-L.;
RT   "Expression and alternative splicing of human factor B gene in
RT   leukemic mononuclear cells.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 20-764, PARTIAL SEQUENCE FROM N.A., AND CARBOHYDRATES.
RX   MEDLINE=84161997; PubMed=6546754;
RA   Mole J.E., Anderson J.K., Davison E.A., Woods D.E.;
RT   "Complete primary structure for the zymogen of human complement
RT   factor B.";
RL   J. Biol. Chem. 259:4407-4412(1984).
RN   [9]
RP   SEQUENCE OF 260-764.
RX   MEDLINE=84204002; PubMed=6342610;
RA   Christie D.,, Gagnon J.;
RT   "Amino acid sequence of the Bb fragment from complement factor B.
```

```
RT   Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and
RT   completion of the sequence of the Bb fragment.";
RL   Biochem. J. 209:61-70(1983).
RN   [10]
RP   SEQUENCE OF 339-764 FROM N.A.
RX   MEDLINE-83273641; PubMed-6308626;
RA   Campbell R.D., Porter R.R.;
RT   "Molecular cloning and characterization of the gene coding for human
RT   complement protein factor B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4464-4468(1983).
RN   [11]
RP   SEQUENCE OF 462-595 AND 752-764 FROM N.A.
RX   MEDLINE-83039428; PubMed-6957884;
RA   Woods D.E., Markham A.F., Ricker A.T., Goldberger G., Colten H.R.;
RT   "Isolation of cDNA clones for the human complement protein factor B,
RT   a class III major histocompatibility complex gene product.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:5661-5665(1982).
RN   [12]
RP   SEQUENCE OF 16-259 FROM N.A.
RX   MEDLINE-84158524; PubMed-6423161;
RA   Morley B.J., Campbell R.D.;
RT   "Internal homologies of the Ba fragment from human complement
RT   component Factor B, a class III MHC antigen.";
RL   EMBO J. 3:153-157(1984).
RN   [13]
RP   SEQUENCE OF 1-99 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE-87102880; PubMed-3643061;
RA   Wu L.C., Morley B.J., Campbell R.D.;
RT   "Cell specific expression of the human complement protein factor B
RT   gene: evidence for the role of two distinct 5'-flanking elements.";
RL   Cell 48:331-342(1987).
RN   [14]
RP   GLYCATION IN POSITION 291.
RX   MEDLINE-91174758; PubMed-2006911;
RA   Niemann M.A., Bhown A.S., Miller E.J.;
RT   "The principal site of glycation of human complement factor B.";
RL   Biochem. J. 274:473-480(1991).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE. IT HAS ALSO BEEN IMPLICATED
CC       IN PROLIFERATION AND DIFFERENTIATION OF PREACTIVATED B
CC       LYMPHOCYTES, RAPID SPREADING OF PERIPHERAL BLOOD MONOCYTES,
CC       STIMULATION OF LYMPHOCYTE BLASTOGENESIS AND LYSIS OF ERYTHROCYTES.
CC       BA INHIBITS THE PROLIFERATION OF PREACTIVATED B LYMPHOCYTES.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha-chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha-chain to yield C5a and C5b. Both
CC       cleavages take place at the C terminal of an arginine residue.
CC   -!- SUBUNIT: MONOMER.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- POLYMORPHISM: TWO MAJOR VARIANTS, F AND S, AND 2 MINOR VARIANTS,
CC       AS WELL AS AT LEAST 14 VERY RARE VARIANTS, HAVE BEEN IDENTIFIED.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72875; CAA51389.1; -.
DR   EMBL; S67310; AAD13989.1; -.
DR   EMBL; L15702; AAA16820.1; -.
DR   EMBL; X00284; CAA25077.1; -.
DR   EMBL; AF019613; AAB67977.1; -.
DR   EMBL; BC004443; AAH04143.1; -.
DR   EMBL; AF349079; AAK30167.1; -.
DR   EMBL; K01566; AAA36225.2; -.
DR   EMBL; J00127; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00126; AAA36226.1; -.
DR   EMBL; J00185; AAA36219.1; ALT_SEQ.
DR   EMBL; J00186; AAA36220.1; -.
DR   EMBL; M15082; AAA59625.1; -.
DR   PIR; A00934; BBHU.
DR   PIR; S14339; S14339.
DR   PIR; S34075; S34075.
DR   HSSP; P00734; 2HNT.
DR   MEROPS; S01.196; -.
DR   SWISS-2DPAGE; P00751; HUMAN.
DR   Siena 2DPAGE; P00751; -.
DR   MIM; 138470; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 3.
DR   Pfam; PF00089; trypsin; 1.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi; Signal; Polymorphism; Zymogen;
KW   Alternative splicing.
FT   SIGNAL        1     25
FT   CHAIN        26    764       COMPLEMENT FACTOR B.
FT   CHAIN        26    259       BA FRAGMENT.
FT   CHAIN       260    764       BB FRAGMENT.
FT   DOMAIN       38     99       SUSHI 1
FT   DOMAIN      104    159       SUSHI 2
FT   DOMAIN      164    219       SUSHI 3
FT   DOMAIN      277    469       VWFA.
FT   DOMAIN      484    764       SERINE PROTEASE.
FT   ACT_SITE    526    526       CHARGE RELAY SYSTEM.
FT   ACT_SITE    575    576       CHARGE RELAY SYSTEM.
FT   ACT_SITE    694    699       CHARGE RELAY SYSTEM.
FT   DISULFID     37     76       BY SIMILARITY.
FT   DISULFID     62     98       BY SIMILARITY.
FT   DISULFID    103    145       BY SIMILARITY.
FT   DISULFID    131    158       BY SIMILARITY.
FT   DISULFID    165    205       BY SIMILARITY.

  Query Match             17.7%;  Score 133;  DB 1;  Length 764;
  Best Local Similarity   29.9%;  Pred. No. 7.1e-05;
  Matches   40;  Conservative   15;  Mismatches   54;  Indels   16;  Gaps    6;

Qy       2 ISCCSPPPTLNC----RISYYSTPIAVGTVIEYSCSGTFRLVGEKSLLCITKDKVDGTWD 57
           | |  |    ||    |  ||:    |   | :     : | |  :  |    :|:| |
Db     101 IHCPRPHDFENGEYWPRSPYYN----VSDEISFHCYDGYTLRGSANRTC----QVNGRWS 152

Qy      58 KPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
             |:  |    |  | :| | :  ||  ||   |||:      ::  ::   ||
Db     153 GQTAICD--NGAGYCSNPGIPIGTRKVGS-QYR-LEDSVTYHCSRGLTLRGSQRRTCQEGG 209

Qy     118 MWGPTRLPTCVSVF 131
           |  |  |:|  |
Db     210 SWSGTE-PSCQDSF 222


RESULT 34
```

```
APOH_CANFA
ID   APOH_CANFA     STANDARD;      PRT:   345 AA.
AC   P33768;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo H) (B2GPI)
DE   (Beta(2)GPI).
GN   APOH
OS   Canis familiaris (dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN BEAGLE; TISSUE=Liver;
RX   MEDLINE=94231500; PubMed=7982067;
RA   Sellar G.C., Keane J., Mehdi H., Peeples M.E., Browne N.,
RA   Whitehead A.S.;
RT   "Characterization and acute phase modulation of canine apolipoprotein
RT   H (beta 2-glycoprotein I).";
RL   Biochem. Biophys. Res. Commun. 191:1288-1293(1993).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72933; CAA51438.1;
DR   PIR; JN0465; JN0465.
DR   HSSP; P02749; 1C1Z.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Ptam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19       BY SIMILARITY.
FT   CHAIN       20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      22     80       SUSHI 1.
FT   DOMAIN      83    138       SUSHI 2.
FT   DOMAIN     141    201       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     263    345       SUSHI-LIKE.
FT   DISULFID    23     66       BY SIMILARITY.
FT   DISULFID    51     79       BY SIMILARITY.
FT   DISULFID    84    124       BY SIMILARITY.
FT   DISULFID   110    137       BY SIMILARITY.
FT   DISULFID   142    188       BY SIMILARITY.
FT   DISULFID   174    200       BY SIMILARITY.
FT   DISULFID   205    248       BY SIMILARITY.
FT   DISULFID   244    260       BY SIMILARITY.
FT   DISULFID   264    315       BY SIMILARITY.
FT   DISULFID   300    325       BY SIMILARITY.
FT   DISULFID   307    345       BY SIMILARITY.
FT   CARBOHYD   117    117       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   183    183       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   193    193       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   345 AA;  38403 MW;  E0B262487B7B74EEA CRC64;

  Query Match             17.5%;  Score 142;  DB 1;  Length 345;
  Best Local Similarity   28.7%;  Prob. No. 3.8e-06;
  Matches   47;  Conservative   16;  Mismatches   56;  Indels   20;  Gaps    6;

Qy      12 NGRISYYSIPIAVGIVIGVSCSCIFRELCRESLCCKHEVEIWRRAPCCEYENEYSS 71
           || : | |       | :::|  | | | |  |       :| |   | |      :
Db      92 NGAVRY--TTFEYPNTISFACNTGFYLNGSSSAKCTE --EGKWSVDLPVCTRV    T 141

Qy      72 TPEPIVP --GGYKIRGSTPYRHGDSVTFACPTNFYMNGNKSVWCANNMWSGTRLPT 126
           || | ||      :|  :   :|:   | |  :::| || :: | |:  |  | ||
Db     142 CPFPSVPRFAKLSVIKRGAINNSLYCNEAVFLCLPHYAMFCNDLSICIAHCNW  TILPE 199

Qy     127 CVSV-  FP 142
           |  |   +
Db     200 CREVKCPFP 208


RESULT 35
APAR_PIG
ID   APAR_PIG       STANDARD;      PRT:   202 AA.
AC   Q03472;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Apolipoprotein R precursor (Apo-R).
GN   APOR.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 29-47.
RC   TISSUE=Liver;
RX   MEDLINE=93099101; PubMed=1463721;
RA   Cooper S.T., Attie A.D.;
RT   "Pig apolipoprotein R: a new member of the short consensus repeat
RT   family of proteins.";
RL   Biochemistry 31:12428-12436(1992).
CC   -!- FUNCTION: MAY BE A LIPOPROTEIN RECEPTOR THAT LIBERATES THE
CC       COAGULATION OR THE COMPLEMENT CASCADES.
CC   -!- SUBUNIT: FORMS HIGH MOLECULAR WEIGHT DISULFIDE-LINKED COMPLEXES.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: PLASMA. FOUND ON VERY LOW DENSITY LIPOPROTEIN
CC       (VLDL), ON CHYLOMICRONS, AND IN THE 1.19-1.21 G/ML FRACTION OF PIG
CC       PLASMA. FOUND IN PIG LIVER, SPLEEN, LUNG, BONE MARROW AND LYMPH
CC       NODE.
CC   -!- DEVELOPMENTAL STAGE: TERMINALLY DIFFERENTIATED MACROPHAGES.
CC   -!- SIMILARITY: WITH THE C-TERMINUS OF HUMAN C4B BINDING PROTEIN.
CC       HOMOLOGY WITH A FAMILY OF PROTEINS INVOLVED IN COAGULATION AND
CC       COMPLEMENT.
CC   -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06820; AAA30994.1;
DR   PIR; JS0773; JS0773.
DR   PIR; A44247; A44247.
DR   HSSP; P10996; 1VVC.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Ptam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Lipid transport; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     28
FT   CHAIN       29    202       APOLIPOPROTEIN R.
FT   DOMAIN      29     86       SUSHI 1.
FT   DOMAIN      88    144       SUSHI 2.
FT   DISULFID    36     73       BY SIMILARITY.
FT   DISULFID    59     85       BY SIMILARITY.
FT   DISULFID    89    130       BY SIMILARITY.
FT   DISULFID   116    143       BY SIMILARITY.
FT   DISULFID   151    151       INTERCHAIN (BY SIMILARITY).
```

```
FT   DISULFID    163    163       INTERCHAIN (BY SIMILARITY).
FT   CONFLICT     44     44       I -> V (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     46     46       G -> A (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   202 AA;  22724 MW;  EABC951B06A0D80C CRC64;

  Query Match             17.4%;  Score 131;  DB 1;  Length 202;
  Best Local Similarity   25.2%;  Pred. No. 2.7e-06;
  Matches   34;  Conservative   23;  Mismatches   56;  Indels   22;  Gaps    6;

Qy       4 CGSPPPILNGRISYYSTPIAVGTV------IRYSCSGTFRLIGEKSLLCITKDKVDGTWD 57
           | :|| : :|    : | | :|        : | |   : |:||  | | :       |
Db      30 CDNPPVVAHG----HHTQI-IGLFGMKKDEVVYKCDEGYTLVGEDRLSCKS-----SRWS 79

Qy      58 KPAPKCEYPNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
             ||:|:     : ||:| :  |        |   ::|   | : : : | | : |  :
Db      80 PAAPQCK-----ALCPKPQIDRGKLSVDQDEYIESENVIVQCGSGYGLVGPKIITCTEDG 134

Qy     118 MWGPTRLPTCVSVFP 132
           | | |:| |   :|
Db     135 TWHP RVPKCEWEYP 148


RESULT 36
CFAH_BOVIN
ID   CFAH_BOVIN     STANDARD;      PRT;   685 AA.
AC   Q28085;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement factor H (H factor 1) (Fragments).
GN   HF1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16
RT   TISSUE=Liver;
RX   MEDLINE=96202005; PubMed=8615824;
RA   Soames C.J., Day A.J., Sim R.B.;
RT   "Prediction from sequence comparisons of residues of factor H involved
RT   in the interaction with complement component C3b.";
RL   Biochem. J. 315:523-531(1996).
CC   -!- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC       C3b by factor I and also increases the rate of dissociation of the
CC       C3bBb complex (C3 convertase) and the (C3b)NBB complex (C5
CC       convertase) in the alternative complement pathway (By similarity).
CC   -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X98697; CAA67257.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 11.
DR   SMART; SM00032; CCP; 11.
KW   Complement alternate pathway; Plasma; Repeat; Sushi.
FT   NON_TER       1      1
FT   NON_CONS     16     17
FT   DOMAIN      <17     67       SUSHI 2.
FT   DOMAIN       70    131       SUSHI 3.
FT   DOMAIN      134    188       SUSHI 4.
FT   DOMAIN      191    246       SUSHI 5.
FT   DOMAIN      249    309       SUSHI 6.
FT   DOMAIN      311    366       SUSHI 7.
FT   DOMAIN      370    429       SUSHI 8.
FT   DOMAIN      431    488       SUSHI 9.
FT   DOMAIN      491    547       SUSHI 10.
FT   DOMAIN      553    609       SUSHI 11.
FT   DOMAIN      614    668       SUSHI 12.
FT   DOMAIN      675   >685       SUSHI 13.
FT   DISULFID     39     66       BY SIMILARITY.
FT   DISULFID     71    117       BY SIMILARITY.
FT   DISULFID    103    130       BY SIMILARITY.
FT   DISULFID    135    176       BY SIMILARITY.
FT   DISULFID    162    187       BY SIMILARITY.
FT   DISULFID    192    234       BY SIMILARITY.
FT   DISULFID    219    245       BY SIMILARITY.
FT   DISULFID    250    297       BY SIMILARITY.
FT   DISULFID    280    308       BY SIMILARITY.
FT   DISULFID    312    354       BY SIMILARITY.
FT   DISULFID    339    365       BY SIMILARITY.
FT   DISULFID    371    417       BY SIMILARITY.
FT   DISULFID    400    428       BY SIMILARITY.
FT   DISULFID    432    476       BY SIMILARITY.
FT   DISULFID    459    487       BY SIMILARITY.
FT   DISULFID    492    534       BY SIMILARITY.
FT   DISULFID    520    546       BY SIMILARITY.
FT   DISULFID    554    597       BY SIMILARITY.
FT   DISULFID    583    608       BY SIMILARITY.
FT   DISULFID    615    656       BY SIMILARITY.
FT   DISULFID    642    667       BY SIMILARITY.
FT   NON_TER     685    685
SQ   SEQUENCE   685 AA;  77536 MW;  69FC9[illegible]C8D530E872 CRC64;

  Query Match             17.4%;  Score 131;  DB 1;  Length 685;
  Best Local Similarity   28.5%;  Pred. No. 1e-05;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   40;  Gaps    8;

Qy       3 SCGSPPPILNG-RISYYSTPIAVGIVIRYSCSGTFRLIGEKSLLCITKDKVDGTWLKPAP 61
           ||  || : || |   :     |   |: :|:|  | : |   : |      :|     |
Db     553 SCALPPELPNGKRKEIHKEEYAHNEVVEYACNPKFLMKGSHKIQC--  --VDGE-SIALP 606

Qy      62 KCEYPNKYSSCP---------PPIVPGGYKFIGSTPYRHGDSVTFACKTNFSMNGNKSVW 112
           |       :|           :( ||          |  |||||  |:|:  :|| | : :
Db     607 VC  IEEERTCEISDLDHGDVKPSVP------  ---PIHHGDSVEFSCKEAFIMIGPRFII 655

Qy     113 CQANNMWGPTRLPTCVS 129
           |  :     |: | |::
Db     656 C-ISGFW--TQPPQCIA 669


RESULT 37
LEM3_SHEEP
ID   LEM3_SHEEP     STANDARD;      PRT;   769 AA.
AC   P98109;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RA   Burns S.A., Neufeld E.J., Donady J.J.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
```

```
CC   !  SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC   !  SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L34270; AAB59261.1; -.
DR   HSSP; P16109; 1FSB.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00344; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     32       POTENTIAL.
FT   CHAIN        33    709       P-SELECTIN.
FT   DOMAIN       33    715       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    716    738       POTENTIAL.
FT   DOMAIN      739    769       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58    156       C-TYPE LECTIN.
FT   DOMAIN      159    199       EGF-LIKE.
FT   DOMAIN      199    258       SUSHI 1.
FT   DOMAIN      261    320       SUSHI 2.
FT   DOMAIN      323    382       SUSHI 3.
FT   DOMAIN      385    444       SUSHI 4.
FT   DOMAIN      447    506       SUSHI 5.
FT   DOMAIN      509    568       SUSHI 6.
FT   DOMAIN      580    639       SUSHI 7.
FT   DOMAIN      642    701       SUSHI 8.
FT   DISULFID     60    156       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    163    174       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    185    194       BY SIMILARITY.
FT   DISULFID    200    244       BY SIMILARITY.
FT   DISULFID    230    257       BY SIMILARITY.
FT   DISULFID    262    306       BY SIMILARITY.
FT   DISULFID    292    319       BY SIMILARITY.
FT   DISULFID    324    368       BY SIMILARITY.
FT   DISULFID    354    381       BY SIMILARITY.
FT   DISULFID    386    430       BY SIMILARITY.
FT   DISULFID    416    443       BY SIMILARITY.
FT   DISULFID    448    492       BY SIMILARITY.
FT   DISULFID    478    505       BY SIMILARITY.
FT   DISULFID    510    554       BY SIMILARITY.
FT   DISULFID    540    567       BY SIMILARITY.
FT   DISULFID    581    625       BY SIMILARITY.
FT   DISULFID    611    638       BY SIMILARITY.
FT   DISULFID    643    687       BY SIMILARITY.
FT   DISULFID    673    700       BY SIMILARITY.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    498    498       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    604    604       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    655    655       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    662    662       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        757    760       ENDOCYTOSIS SIGNAL (PROBABLE).
FT   VARIANT     566    566       S -> T.
FT   VARIANT     579    579       I -> V.
SQ   SEQUENCE   769 AA;  84317 MW;  23E42575D60FAB15 CRC64;

  Query Match             17.4%;  Score 131;  DB 1;  Length 769;
  Best Local Similarity   29.1%;  Pred. No. 1.1e-05;
  Matches   89;  Conservative   18;  Mismatches   59;  Indels   18;  Gaps    7;

Qy       1 GISCGSPPEILNGRIS--YYSIPEAVGIVIFVSCIFLIVIVIAIWR 58
           |: | |       | :|  ::     : |   : |   | |:|| : | |     :   |
Db     578 GLRCPSLIAPNQGTMSCQHHLRNFGLNFTCHFGCKAGFTLMGESALQC  --RPSRQWTA 633

Qy      59 PAPPCPYPNYYSCPI--FPIVPLCYPIWCFP   YPPCIYVIACPINLSMNGRSVWC 114
           |:|     | | |  |||     :  | |   : :| : :| |     :|||:   |
Db     634 VAPTCRAV-KCSKLPVTEPIV-  --MNCSNPWGNFSYGSTCSFPCPEQLLNGSERTAC 687

Qy     114 QANNMWGPTRLPTC 127
           | |  |  | :|||
Db     688 QENGQWS-TTMPTC 700


RESULT 38
LEM2_RAT
ID   LEM2_RAT        STANDARD;       PRT:   549 AA.
AC   P98105;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE OR ELAM-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Rosenbloom C.L., Auchampach J.A., Anderson D.C., Manning A.M.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L25527; AAA41113.1; -.
DR   HSSP; P16581; 1KJA.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
```

```
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    549       E-SELECTIN.
FT   DOMAIN       22    494       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    495    516       POTENTIAL.
FT   DOMAIN      517    549       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       38    138       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      139    175       EGF LIKE.
FT   DOMAIN      179    239       SUSHI 1.
FT   DOMAIN      242    301       SUSHI 2.
FT   DOMAIN      304    364       SUSHI 3.
FT   DOMAIN      367    427       SUSHI 4.
FT   DOMAIN      430    486       SUSHI 5.
FT   DISULFID     40    138       BY SIMILARITY.
FT   DISULFID    111    130       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    163       BY SIMILARITY.
FT   DISULFID    165    174       BY SIMILARITY.
FT   DISULFID    180    225       BY SIMILARITY
FT   DISULFID    210    238       BY SIMILARITY.
FT   DISULFID    243    287       BY SIMILARITY.
FT   DISULFID    273    300       BY SIMILARITY
FT   DISULFID    305    350       BY SIMILARITY.
FT   DISULFID    336    363       BY SIMILARITY.
FT   DISULFID    368    413       BY SIMILARITY
FT   DISULFID    399    426       BY SIMILARITY.
FT   DISULFID    441    472       BY SIMILARITY.
FT   DISULFID    458    485       BY SIMILARITY.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD    192    192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD    313    313       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD    320    320       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    441    441       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD    465    465       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   549 AA;  60079 MW;  85CEECDB7B0144C8 CRC64,

  Query Match             17.2%;  Score 129.5;  DB 1;  Length 549;
  Best Local Similarity   29.4%;  Pred. No. 1.1e-05;
  Matches   32;  Conservative   12;  Mismatches   56;  Indels    9;  Gaps    3;

Qy      19 STPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAPKCEYFNKYSSCPEPIVP 78
           |  |   |:  .:||   | | | : | |  :    | ||   | |         |    :|
Db     385 SAPFQSGSSCKFSCDEGFELKGSRRLQCGPR---GEWDSEKPTCAGV---QCSSLDLP 437

Qy      79 GGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPTRLPTC 127
             |   : | |   |    | |   :::||:  : | |   |    ||||
Db     438 GKMNMSCSGPAVFGTVCEFTCPEGWTLNGSSILTCGATGRWS-AMLPTC 485


RESULT 37
APOH_BOVIN
ID   APOH_BOVIN     STANDARD;      PRT;   345 AA.
AC   P17690; Q28052;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI).
GN   APOH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver
RA   Gao B., Vitmani M., Ronni E., Lazar-Wesley E., Sakaguchi K.,
RA   Appella E., Kamos G., Takacs L.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 4-345 FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RC   TISSUE=Liver
RX   MEDLINE=92232647; PubMed=1567819;
RA   Bendixen E., Halkier T., Magnusson S., Sottrup-Jensen L.,
RA   Kristensen T.;
RT   "Complete primary structure of bovine beta 2-glycoprotein I:
RT   localization of the disulfide bridges.";
RL   Biochemistry 31:3611-3617(1992).
RN   [3]
RP   SEQUENCE OF 20-345, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RC   TISSUE=Plasma;
RX   MEDLINE=92089075; PubMed=1751487;
RA   Kato H., Enjyoji K.-I.;
RT   "Amino acid sequence and location of the disulfide bonds in bovine
RT   beta 2 glycoprotein I: the presence of five Sushi domains.";
RL   Biochemistry 30:11687-11694(1991).
RN   [4]
RP   SEQUENCE OF 20-41.
RX   MEDLINE=90226328; PubMed=2327984;
RA   Li C., Blacher R., Esch F., Congote L.F.;
RT   "Isolation from fetal bovine serum of an apolipoprotein-H like
RT   protein which inhibits thymidine incorporation in fetal calf
RT   erythroid cells.";
RL   Biochem. J. 267:261-264(1990).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ----------------------------------------------------------------------
DR   EMBL; L07303; AAA30382.1; -.
DR   EMBL; X60065; CAA42669.1; -.
DR   PIR; S23597; NIBO.
DR   PIR; S09032; S09032.
DR   HSSP; P02749; 1CLZ.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN       22     80       SUSHI 1.
FT   DOMAIN       83    138       SUSHI 2.
FT   DOMAIN      141    201       SUSHI 3.
FT   DOMAIN      204    261       SUSHI 4.
FT   DOMAIN      263    345       SUSHI-LIKE.
FT   DISULFID     23     66
FT   DISULFID     51     79
FT   DISULFID     84    124
FT   DISULFID    110    137
FT   DISULFID    142    188
FT   DISULFID    174    200
FT   DISULFID    205    248
```

```
FT   DISULFID     234     260
FT   DISULFID     294     315
FT   DISULFID     803     825
FT   DISULFID     807     845
FT   CARBOHYD      92      92       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     162     162       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     183     183       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     193     193       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     253     253       N-LINKED (GLCNAC. . .).
FT   CONFLICT     101     101       E   -> G (IN REF. 1).
FT   CONFLICT     108     108       F   -> S (IN REF. 1).
FT   CONFLICT     177     177       H   -> R (IN REF. 3).
FT   CONFLICT     191     191       H   -> N (IN REF. 3).
FT   CONFLICT     194     194       W   -> C (IN REF. 1).
FT   CONFLICT     259     259       S   -> N (IN REF. 1).
FT   CONFLICT     302     302       H   -> N (IN REF. 1).
FT   CONFLICT     305     305       K   -> R (IN REF. 1).
FT   CONFLICT     329     329       H   -> R (IN REF. 1).
SQ   SEQUENCE   345 AA;  38252 MW;  F1170AB6094A1C33 CRC64;

  Query Match          16.9%;  Score 127;  DB 1;  Length 345;
  Best Local Similarity   30.0%;  Pred. No. 1.2e-05;
  Matches   49;  Conservative   14;  Mismatches   55;  Indels   22;  Gaps    8;

Qy       12 NGRISYYSTFTIAVGIVIRYSGSGIEKKIGEKSIAYTLKGKVDGTWDKPAPKCEYFNKYSS 71
            ||  :  |  |     |  :||   | | |  |  |       :| |    | |     :
Db       62 NGIVRY  TTFEYFNTISFSCHTGFYLKGASSAKCTE    EGKWSPDLPVCAPI-  T 141

Qy       72 CPEPIVP      GGYK IRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPTRLP 125
            || | :|      || : |:  : :|    | |  : :| || :| |  :  |  |:||
Db      142 CPFPPTLKFASLSVYKPLAGNNSF YGSKAVFKCLPHRAMFGNDTVTCTEHGNW  FQLP 198

Qy      126 ICVSV   FP 142
              |   ||
Db      199 ECREVRCPFP 208


RESULT 13
APOH_HUMAN
ID   APOH_HUMAN      STANDARD;      PRT;   345 AA.
AC   P02749;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo H) (B2GPI)
DE   (Beta(2)GPI) (Activated protein C binding protein) (APC inhibitor).
GN   APOH OR B2GI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91315408; PubMed=1650181;
RA   Steinkasserer A., Estaller C., Weiss E., Sim R.B., Day A.J.;
RT   "Complete nucleotide and deduced amino acid sequence of human beta 2-
RT   glycoprotein I.";
RL   Biochem. J. 277:387-391(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92008618; PubMed=1655524;
RA   Kristensen T., Schousboe I., Boel E., Mulvihill E.M., Hansen R.R.,
RA   Moeller K.B., Moeller N.P.H., Sottrup-Jensen L.;
RT   "Molecular cloning and mammalian expression of human beta
RT   2-glycoprotein I cDNA.";
RL   FEBS Lett. 289:183-186(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92084151; PubMed=1748314;
RA   Mehdi H., Nunn M., Steel D.M., Whitehead A.S., Perez M., Walker I.,
RA   Peeples M.E.;
RT   "Nucleotide sequence and expression of the human gene encoding
RT   apolipoprotein H (beta 2-glycoprotein I).";
RL   Gene 108:293-298(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92273779; PubMed=1339416;
RA   Day J.R., O'Hara P.J., Grant F.J., Lofton-Day C.E., Beckaw M.N.,
RA   Werner P., Arnaud P.;
RT   "Molecular cloning and sequence analysis of the cDNA encoding human
RT   apolipoprotein H (beta 2-glycoprotein I).";
RL   Int. J. Clin. Lab. Res. 21:256-264(1992).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92135065; PubMed=1777418;
RA   Matsuura E., Igarashi M., Igarashi Y., Narao H., Ichikawa K.,
RA   Yasuda T., Koike T.;
RT   "Molecular definition of human beta 2-glycoprotein I (beta 2-GPI) by
RT   cDNA cloning and inter-species differences of beta 2-GPI in
RT   alternation of anticardiolipin binding.";
RL   Int. Immunol. 3:1217-1221(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99115472; PubMed=9914524;
RA   Okkels H., Rasmussen I.E., Sanghera D.K., Kamboh M.I., Kristensen T.;
RT   "Structure of the human beta2-glycoprotein I (apolipoprotein H)
RT   gene.";
RL   Eur. J. Biochem. 259:435-440(1999).
RN   [7]
RP   SEQUENCE OF 20-345, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RX   MEDLINE=84222015; PubMed=6587378;
RA   Lozier J., Takahashi N., Putnam F.W.;
RT   "Complete amino acid sequence of human plasma beta 2-glycoprotein I.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:3640-3644(1984).
RN   [8]
RP   SEQUENCE OF 20-38.
RC   TISSUE=Follicular fluid;
RX   MEDLINE=21248139; PubMed=11250549;
RA   Aleporou-Marinou V., Pappa H., Yalouris P., Patarotas T.;
RT   "Purification of apolipoprotein H (beta 2-glycoprotein I) like protein
RT   from human follicular fluid.";
RL   Comp. Biochem. Physiol. 128B:537-542(2001).
RN   [9]
RP   DISULFIDE BONDS IN C-TERMINAL DOMAIN.
RX   MEDLINE=93050249; PubMed=1426288;
RA   Steinkasserer A., Barlow P.N., Willis A.C., Kertesz Z.,
RA   Campbell I.D., Sim R.B., Norman D.G.;
RT   "Activity, disulphide mapping and structural modelling of the fifth
RT   domain of human beta 2-glycoprotein I.";
RL   FEBS Lett. 323:194-197(1992).
RN   [10]
RP   STRUCTURE OF CARBOHYDRATES.
RX   MEDLINE=97299942; PubMed=9155091;
RA   Gambino R., Ruin G., Pagano G., Cassador M.;
RT   "Qualitative analysis of the carbohydrate composition of
RT   apolipoprotein H.";
RL   J. Protein Chem. 16:205-212(1997).
RN   [11]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RC   TISSUE=Plasma;
RX   MEDLINE=99437994; PubMed=10508150;
RA   Bouma B., de Groot P.G., van Den Elsen J.M.H., Ravelli R.B.G.,
RA   Schouten A., Simmelink M.J.A., Derksen R.H.W.M., Kroon J., Gros P.;
RT   "Adhesion mechanism of human beta(2)-glycoprotein I to phospholipids
RT   based on its crystal structure.";
RL   EMBO J. 18:5166-5174(1999).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.87 ANGSTROMS).
RX   MEDLINE=20031634; PubMed=10562545;
RA   Schwarzenbacher R., Zeth K., Diederichs K., Gries A., Kostner G.M.,
RA   Laggner P., Prassl R.;
RT   "Crystal structure of human beta2-glycoprotein I: implications for
RT   phospholipid binding and the antiphospholipid syndrome.";
```

```
RL   EMBO J. 18:6228-6239(1999).
RN   [13]
RP   VARIANT LEU 266.
RX   MEDLINE-93273313; PubMed-8099061;
RA   Steinkasserer A., Doerner C., Wherzner R., Sim R.B.;
RT   "Human beta 2-glycoprotein I: molecular analysis of DNA and amino
RT   acid polymorphism.";
RL   Hum. Genet. 91:401-402(1993).
RN   [14]
RP   VARIANT ASN-107.
RX   MEDLINE-97369481; PubMed-9225969;
RA   Sanghera D.K., Kristensen T., Hamman R.F., Kamboh M.I.;
RT   "Molecular basis of the apolipoprotein H (beta 2-glycoprotein I)
RT   protein polymorphism.";
RL   Hum. Genet. 100:57-62(1997).
RN   [15]
RP   VARIANTS GLY 325 AND SER-335.
RX   MEDLINE-97217791; PubMed-9063752;
RA   Sanghera D.K., Wagenknecht D.R., McIntyre J.A., Kamboh M.I.;
RT   "Identification of structural mutations in the fifth domain of
RT   apolipoprotein H (beta-2-glycoprotein I) which affect phospholipid
RT   binding.";
RL   Hum. Mol. Genet. 6:411-416(1997).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X58100; CAA41114.1; -.
DR   EMBL; X53595; CAA37664.1; -.
DR   EMBL; X57847; CAA40977.1; -.
DR   EMBL; M62839; AAA51766.1; -.
DR   EMBL; S80305; AAB21340.1; -.
DR   EMBL; Y11493; CAA72279.1; -.
DR   EMBL; Y11494; CAA72279.1; JOINED.
DR   EMBL; Y11495; CAA72279.1; JOINED.
DR   EMBL; X53595; CAA72279.1; JOINED.
DR   EMBL; Y11496; CAA72279.1; JOINED.
DR   EMBL; Y11497; CAA72279.1; JOINED.
DR   EMBL; Y11498; CAA72279.1; JOINED.
DR   EMBL; Y17754; CAA76845.1;
DR   PIR; S17178; NBHU.
DR   PIR; B43286; B43286.
DR   PDB; 1QUB; 08-OCT-99.
DR   PDB; 1C1Z; 19-NOV-99.
DR   MIM; 138700; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal;
KW   Polymorphism; 3D-structure.
FT   SIGNAL       1     19
FT   CHAIN       20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      22     80       SUSHI 1.
FT   DOMAIN      83    140       SUSHI 2.
FT   DOMAIN     141    201       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     263    345       SUSHI-LIKE.
FT   DISULFID    23     66
FT   DISULFID    51     79
FT   DISULFID    84    124
FT   DISULFID   110    147
FT   DISULFID   142    188
FT   DISULFID   174    200
FT   DISULFID   205    248
FT   DISULFID   234    260
FT   DISULFID   264    315
FT   DISULFID   300    325
FT   DISULFID   307    345
FT   CARBOHYD   149    149       O-LINKED.
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...).
FT   CARBOHYD   183    183       N-LINKED (GLCNAC...).
FT   CARBOHYD   193    193       N-LINKED (GLCNAC...).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC...).
FT   VARIANT    107    107       S -> N (N APOH*1).
FT                               /FTId=VAR_008169.
FT   VARIANT    266    266       V -> L (N 23% OF THE POPULATION).
FT                               /FTId=VAR_000673.
FT   VARIANT    325    325       G -> S (LOSS OF PHOSPHATIDYLSERINE
FT                               BINDING
FT                               /FTId=VAR_008170.
FT   VARIANT    335    335       W -> S (N APOH*4W; LOSS OF
FT                               PHOSPHATIDYLSERINE-BINDING)

  Query Match          16.9%;  Score 127;  DB 1;  Length 345;
  Best Local Similarity   29.8%;  Pred. No. 1.2e-05;
  Matches   39;  Conservative   12;  Mismatches   56;  Indels   24;  Gaps    7;

Qy      12 NGRISYYSTPIAVGTVIRYSCSGTFRLVGEKSLLCITKDKVDGTWDKPAPKCEYFNKYSS 71
           || : |  |       | :||:  | |    |  |       :| |    | |
Db      92 NGAVRY--TTFEYPNTISFSCNTGFYLNGADSAKCTE---EGKWSPELPVCAPI---I 141

Qy      72 CPEPIVPGGYKIRGSTP------YRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPTRL 124
           || | :    :|  |       ||  |:  | |    :| || :: |  :  |  |:
Db     142 CPPPSIPTFATLRVYKPSAGNNSLYR--DIAVFECLEQHAMPGNDIITCTTHGNW-IKL 197

Qy     125 PTCVSV---FP 132
           | |  |   ||
Db     198 PECKEVKCPFP 208


RESULT 41
LEM3_BOVIN
ID   LEM3_BOVIN     STANDARD;      PRT;   646 AA.
AC   P42201;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Capillary endothelium;
RX   MEDLINE=93249394; PubMed=7683458;
RA   Strubel N.A., Nguyen M., Kansas G.S., Tedder T.F., Bischoff J.;
RT   "Isolation and characterization of a bovine cDNA encoding a
RT   functional homolog of human P-selectin.";
RL   Biochem. Biophys. Res. Commun. 192:338-344(1993).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC       AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CEL-
CC       ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC       THE CELL SURFACE.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE (CALCIUM-DEPENDENT) LECTIN DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
```

```
CC   !  SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS; BOVINE P-LECTIN LACKS
CC      THE HUMAN SUSHI 3, 4 AND 7 EQUIVALENTS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L12041; AAA30743.1; -.
DR   HSSP; P16109; 1ESM.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00348; SELECTIN.
DR   SMART; SM00032; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     41       BY SIMILARITY.
FT   CHAIN        42    646       P-SELECTIN.
FT   DOMAIN       42    587       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    588    611       POTENTIAL.
FT   DOMAIN       612    646       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN       159    195       EGF-LIKE.
FT   DOMAIN       199    258       SUSHI 1.
FT   DOMAIN       261    320       SUSHI 2.
FT   DOMAIN       323    382       SUSHI 3.
FT   DOMAIN       385    444       SUSHI 4.
FT   DOMAIN       457    518       SUSHI 5.
FT   DOMAIN       519    578       SUSHI 6.
FT   DISULFID      60    158       BY SIMILARITY.
FT   DISULFID     141    150       BY SIMILARITY.
FT   DISULFID     163    174       BY SIMILARITY.
FT   DISULFID     168    183       BY SIMILARITY.
FT   DISULFID     185    194       BY SIMILARITY.
FT   DISULFID     200    244       BY SIMILARITY.
FT   DISULFID     240    257       BY SIMILARITY.
FT   DISULFID     262    306       BY SIMILARITY.
FT   DISULFID     292    319       BY SIMILARITY.
FT   DISULFID     324    368       BY SIMILARITY.
FT   DISULFID     354    381       BY SIMILARITY.
FT   DISULFID     386    430       BY SIMILARITY.
FT   DISULFID     416    443       BY SIMILARITY.
FT   DISULFID     458    502       BY SIMILARITY.
FT   DISULFID     488    515       BY SIMILARITY.
FT   DISULFID     520    564       BY SIMILARITY.
FT   DISULFID     550    577       BY SIMILARITY.
FT   CARBOHYD      48     48       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      80     80       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     180    180       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     212    212       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     219    219       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     336    336       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     481    481       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     542    542       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     589    589       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     597    597       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE         634    637       ENDOCYTOSIS SIGNAL (PROBABLE).
SQ   SEQUENCE   646 AA;  71229 MW;  573912A4627A6ACA CRC64;
```

```
  Query Match             16.9%;  Score 127;  DB 1;  Length 646;
  Best Local Similarity   26.9%;  Pred. No. 2.4e-05;
  Matches   36;  Conservative   21;  Mismatches   59;  Indels   18;  Gaps    7;

Qy       1 GISCGSPPPPILNGRIS  YYSIPIAVGIVIFVSCSCIFFLIGKRSLICLIKDKVNIWDK 58
           :: | :    | :|  ::   : |  : :|  | ||:|: :| |   :   |
Db     455 GVQCPTELIAPKQGTMSCQHHVRNFGLNITCHFGCKAGFTLLGDSALQC - RPSRQWTA 510

Qy      59 PAPECEYFNFYSSCP--FFIVFPGYFIDCYTF -YRFCLCVIIACFINFSMNGNKSVWC 113
           || |    | :  |  |||     :  | |   : :| : :| |     ::|::   |
Db     511 AAPICRAV KCAKLPVTEPIV - - MNCSNPWGNFSYGSTCSFHCPEGQLLNGSERIVC 564

Qy     114 QANNMWGPIRLPIC 127
           | |  |  | :||
Db     565 QENGQWS-TTMPTC 577


RESULT 42
LEM2_MOUSE
ID   LEM2_MOUSE      STANDARD;       PRT;   612 AA.
AC   Q00690;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-FEB-1996 (Rel. 34, Last annotation update)
DE   E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE OR ELAM-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92283265; PubMed=1375914;
RA   Hecker-Andre M., van Buijsduijnen R.B., Losberger C., Whelan J.,
RA   Delamarter J.F.;
RT   "Murine endothelial leukocyte adhesion molecule 1 is a close
RT   structural and functional homologue of the human protein.";
RL   Eur. J. Biochem. 206:401-411(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340571; PubMed=1378846;
RA   Weller A., Isenmann S., Vestweber D.;
RT   "Cloning of the mouse endothelial selectins. Expression of both E
RT   and P-selectin is inducible by tumor necrosis factor alpha.";
RL   J. Biol. Chem. 267:15176-15183(1992).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY 2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M80778; AAA37547.1; -.
DR   EMBL; M87862; AAA37577.1; ALT_INIT.
DR   HSSP; P16581; 1KJA.
DR   MGD; MGI:98278; Sele.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
```

```
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00042; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain, Lectin,
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     21       BY SIMILARITY.
FT   CHAIN        22    612       E-SELECTIN.
FT   DOMAIN       22    557       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    558    579       POTENTIAL.
FT   DOMAIN      580    612       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       38    138       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      139    175       EGF-LIKE.
FT   DOMAIN      179    239       SUSHI 1.
FT   DOMAIN      242    301       SUSHI 2.
FT   DOMAIN      304    364       SUSHI 3.
FT   DOMAIN      367    427       SUSHI 4.
FT   DOMAIN      430    490       SUSHI 5.
FT   DOMAIN      493    549       SUSHI 6.
FT   DISULFID     40    138       BY SIMILARITY.
FT   DISULFID    111    130       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    163       BY SIMILARITY.
FT   DISULFID    165    174       BY SIMILARITY.
FT   DISULFID    180    225       BY SIMILARITY.
FT   DISULFID    210    238       BY SIMILARITY.
FT   DISULFID    243    287       BY SIMILARITY.
FT   DISULFID    273    300       BY SIMILARITY.
FT   DISULFID    305    350       BY SIMILARITY.
FT   DISULFID    336    363       BY SIMILARITY.
FT   DISULFID    368    413       BY SIMILARITY.
FT   DISULFID    399    426       BY SIMILARITY.
FT   DISULFID    431    476       BY SIMILARITY.
FT   DISULFID    462    489       BY SIMILARITY.
FT   DISULFID    494    535       BY SIMILARITY.
FT   DISULFID    521    548       BY SIMILARITY.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    192    192       N-LINKED (GLCNAC     ) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    391    391       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    504    504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   612 AA;  66749 MW;  86F85713E9EC2C4D CRC64;

  Query Match             16.4%;  Score 123.5;  DB 1;  Length 612;
  Best Local Similarity   26.5%;  Pred. No. 5.1e-05;
  Matches   36;  Conservative   23;  Mismatches   61;  Indels   16;  Caps    7;

Qy       2 ISCGSPPPILNGRISYYST-PIAVGTVIRYSCSGIFRLIGEKSLLCIIKDKVDGTWDK 58
           ::| : |   || :|  :||   :|  :   ::|  :| | |   : |       | |
Db     303 VTCDAIPQPQNGFVSCSHSTAGELAFKSSCNFTCEQSFTLQGPAQVEC----SAQGQWTP 358

Qy      59 PAPKCEYFNKYSSCPEPIVP--GGYKI--RGSTPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
             | |    |   |      | | |      | |:::| |  |:|:  | : |.. : |
Db     359 QIPVC----KAVQCEALSAPQQGNMKCLPSASGPFQNGSSCEFSCEEGFELKGSRRLQCG 414

Qy     115 ANNMWGPTRLPTCVSV 130
             |  :: ||| :|
Db     415 PRGEW-DSKKPTCSAV 429
```

```
RESULT 43
CO6_HUMAN
ID   CO6_HUMAN       STANDARD;       PRT;   934 AA.
AC   P13671;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement component C6 precursor.
GN   C6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 22-35 AND 633-640.
RX   MEDLINE=90036879; PubMed=2808363;
RA   Haefliger J.-A., Tschopp J., Vial N.; Jeune D.E.;
RT   "Complete primary structure and functional characterization of the
RT   sixth component of the human complement system. Identification of the
RT   C5b-binding domain in complement C6.";
RL   J. Biol. Chem. 264:18041-18051(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89380223; PubMed=2789218;
RA   Discipio R.G., Huqli T.E.;
RT   "The molecular architecture of human complement component C6.";
RL   J. Biol. Chem. 264.16197-16206(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93291175; PubMed=8512929;
RA   Bobart M.J., Fernie B., Discipio R.G.;
RT   "Structure of the human C6 gene.";
RL   Biochemistry 32:6198-6205(1993).
RN   [4]
RP   SEQUENCE OF 1-491 FROM N.A.
RX   MEDLINE=89202413; PubMed=2468158;
RA   Chakravarti D.N., Chakravarti B., Parra C.A., Mueller-Eberhard H.J.;
RT   "Structural homology of complement protein C6 with other
RT   channel-forming proteins of complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2799-2803(1989).
RN   [5]
RP   VARIANT ALLOTYPES C6 A AND C6 B.
RX   MEDLINE=93326158; PubMed=8101442;
RA   Dewald G., Nothen M.M., Cichon S.;
RT   "Polymorphism of human complement component C6: an amino acid
RT   substitution (Glu/Ala) within the second thrombospondin repeat
RT   differentiates between the two common allotypes C6 A and C6 B.";
RL   Biochem. Biophys. Res. Commun. 194:458-464(1993).
RN   [6]
RP   CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=20020247; PubMed=10551839;
RA   Hofsteenge J., Blommers M., Hess D., Furmanek A., Miroshnichenko O.;
RT   "The four terminal components of the complement system are
FT   C-mannosylated on multiple tryptophan residues.";
RL   J. Biol. Chem. 274:32786-32794(1999).
CC   -!- FUNCTION: INVOLVED IN THE FORMATION OF THE LYTIC C5B-9M COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PTM: ALL CYSTEINE RESIDUES ARE ASSUMED TO BE CROSS-LINKED TO ONE
CC       ANOTHER. INDIVIDUAL MODULES CONTAINING AN EVEN NUMBER OF CONSERVED
CC       CYSTEINE RESIDUES ARE SUPPOSED TO HAVE DISULFIDE LINKAGES ONLY
CC       WITHIN THE SAME MODULE.
CC   -!- POLYMORPHISM: The sequence shown is that of allotype C6 A.
CC   -!- SIMILARITY: TO COMPLEMENT FACTORS C7, C8, C9, AND TO PERFORIN.
CC   -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC   ---------- ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

DR EMBL; J05024; AAA51868.1; -.
DR EMBL; J05024; AAA59668.1; -.
DR EMBL; X72177; CAA50994.1; -.
DR EMBL; J04556; AAB59433.1; -.
DR PIR; A32109; A32109.
DR PIR; A34235; A34235.
DR PIR; A34472; A34472.
DR HSSP; P01130; 1AJJ.
DR MIM; 213050; -.
DR InterPro; IPR000561; EGF_like.
DR InterPro; IPR003884; Fact_MAC.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001862; MAC_perforin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000884; TSP.
DR InterPro; IPR002350; kazal.
DR Pfam; PF00050; kazal; 1.
DR Pfam; PF00057; ldl_recept_a; 1.
DR Pfam; PF01823; MACPF; 1.
DR Pfam; PF00084; sushi; 2.
DR Pfam; PF00090; tsp_1; 3.
DR PRINTS; PR00764; COMPLEMENT.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00057; FIMAC; 2.
DR SMART; SM00280; KAZAL; 1.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00457; MACPF; 1.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS00279; MAC_PERFORIN; 1.
DR PROSITE; PS50092; TSP1; 3.
KW Complement pathway; Glycoprotein; Plasma; Membrane attack complex;
KW Cytolysis; Sushi; Repeat; Signal; Polymorphism; EGF-like domain;
KW Transmembrane.
FT SIGNAL 1 21
FT CHAIN 22 934 COMPLEMENT COMPONENT C6.
FT DOMAIN 22 80 TSP TYPE 1 1.
FT DOMAIN 81 137 TSP TYPE 1 2.
FT DOMAIN 138 175 LDL RECEPTOR CLASS A.
FT TRANSMEM 841 849 POTENTIAL.
FT TRANSMEM 854 878 POTENTIAL.
FT DOMAIN 517 558 EGF-LIKE.
FT DOMAIN 562 611 TSP TYPE 1 3.
FT DOMAIN 642 934 C6/C8 BINDING DOMAIN.
FT DOMAIN 643 700 SUSHI 1.
FT DOMAIN 703 762 SUSHI 2.
FT DOMAIN 766 840 COMPLEMENT CONTROL FACTOR 1 MODULE 1.
FT DOMAIN 858 934 COMPLEMENT CONTROL FACTOR 1 MODULE 2.
FT DISULFID 140 151 BY SIMILARITY.
FT DISULFID 146 164 BY SIMILARITY.
FT DISULFID 158 173 BY SIMILARITY
FT DISULFID 399 420 BY SIMILARITY.
FT DISULFID 523 539 BY SIMILARITY.
FT DISULFID 526 541 BY SIMILARITY.
FT DISULFID 543 552 BY SIMILARITY.
FT DISULFID 644 686 BY SIMILARITY.
FT DISULFID 672 699 BY SIMILARITY.
FT DISULFID 704 746 BY SIMILARITY.
FT DISULFID 742 761 BY SIMILARITY.
FT CARBOHYD 29 29 O-LINKED (MAN).
FT CARBOHYD 32 32 O-LINKED (MAN) (PARTIAL).
FT CARBOHYD 90 90 O-LINKED (MAN) (PARTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 568 568 O-LINKED (MAN) (PARTIAL).
FT CARBOHYD 571 571 O-LINKED (MAN) (PARTIAL).
FT CARBOHYD 574 574 O-LINKED (MAN) (PARTIAL).
FT CARBOHYD 855 855 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARIANT [illegible]
FT /FTId=VAR_000056.
SQ SEQUENCE 934 AA; 104843 MW; 5A2713AB5B507B80 CRC64;

Query Match 16.4%; Score 123; DB 1; Length 934;
Best Local Similarity 28.7%; Pred. No. 9.1e-05;
Matches 35; Conservative 16; Mismatches 55; Indels 10; Gaps 5;

Qy 4 CGSPPPIINGRISYYSIPIAVGIVIFVVSCSGIFFCGFFSFIGKRSIDDYRQVPSTWDPIAPPC 63
Db 644 CPQPVPPENGFIRNEKQLYILVGEDVEISCLTGFETVGYQYFRCLP-DGIWRQGDVEC 699

Qy 64 EYFNKYSSCPEPIVPGGYKIRGSTP----YRHGDSVIFACKTNFSMNGNKSVWCQANNMW 119
Db 700 Q--PIECIKPVV---QEVLTITPFQRLYRIGESIELTCPKGFVVAGPSRYTCQGNS W 754

Qy 120 GP 121
Db 752 TP 753

RESULT 44
LEM2_CANFA
ID LEM2_CANFA STANDARD; PRT; 611 AA.
AC P43730;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM 1) (leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Jugular vein;
RA Manning A.M., Lane C.L., Auchampach J.A., Kukielka G.L.,
RA Rosenbloom C.L., Anderson D.C.;
RL Submitted (XXX 1993) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC ---------- [illegible]

CC ---------- [illegible]
DR EMBL; L23087; AAA30843.1; -.
DR HSSP; P16581; 1ESL.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 6.
DR PRINTS; PR00344; SELECTIN.
DR SMART; SM00032; CCP; 6.

```
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL       1     22       BY SIMILARITY.
FT   CHAIN       23    611       E-SELECTIN.
FT   DOMAIN      23    557       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    558   579       POTENTIAL.
FT   DOMAIN      580   611       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39   139       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      140   176       EGF-LIKE.
FT   DOMAIN      180   239       SUSHI 1.
FT   DOMAIN      242   301       SUSHI 2.
FT   DOMAIN      304   364       SUSHI 3.
FT   DOMAIN      367   427       SUSHI 4.
FT   DOMAIN      430   490       SUSHI 5.
FT   DOMAIN      493   549       SUSHI 6.
FT   DISULFID     41   139       BY SIMILARITY.
FT   DISULFID    112   131       BY SIMILARITY.
FT   DISULFID    144   155       BY SIMILARITY.
FT   DISULFID    149   164       BY SIMILARITY.
FT   DISULFID    166   175       BY SIMILARITY
FT   DISULFID    181   225       BY SIMILARITY.
FT   DISULFID    211   238       BY SIMILARITY.
FT   DISULFID    243   287       BY SIMILARITY.
FT   DISULFID    273   300       BY SIMILARITY.
FT   DISULFID    305   350       BY SIMILARITY.
FT   DISULFID    336   363       BY SIMILARITY.
FT   DISULFID    368   413       BY SIMILARITY.
FT   DISULFID    399   426       BY SIMILARITY.
FT   DISULFID    431   476       BY SIMILARITY.
FT   DISULFID    462   489       BY SIMILARITY.
FT   DISULFID    494   535       BY SIMILARITY.
FT   DISULFID    521   548       BY SIMILARITY.
FT   CARBOHYD     26    26       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161   161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    204   204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    266   266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413   413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    433   433       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528   528       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   611 AA;  66414 MW;  45DA9E3DF225E4F6 CRC64;

  Query Match          16.2%;  Score 122;  DB 1;  Length 611;
  Best Local Similarity   25.5%;  Pred. No. 7.2e-05;
  Matches   46;  Conservative   24;  Mismatches   55;  Indels   26;  Gaps    7;

Qy       2 ISCCSPPPTLNGRISYYSTP---IAVGTVIEYSCSGTFRLIGEKSLICITKDKVDGTWDK 58
           ::||:     || :|   :|    :| :   ::|:  | : |   : | :    | | :
Db     304 VTCGAIGHPQNGSVSCSHSPAGEFSVRSSCNFTCNEGFLMQGPAQIECTAQ----GQWSQ 358

Qy      59 PAPKCEYFNKYSSC---PE---PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNK 109
             | |    | | |     ||     :||      |  :: | |  | |:  | : |:|
Db     359 QVPVC----KASQCKALSSPERGYMSCLPG----ASGSFQSGSSCEFFCEKGFVLKGSK 409

Qy     110 SVWCQANNMWGPTRLPTCVSV 130
           :: |     |  :  ||| :|
Db     410 TLQCGLTGKWDSEEPTCEAV 429


RESULT 45
CFAB_MOUSE
ID   CFAB_MOUSE      STANDARD;       PRT;   761 AA.
AC   P04186;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B precursor (EC 3.4.21.47) (C3/C5 convertase).
GN   BF OR B2-BF.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE-91035430; PubMed-2229060;
RA   Ishikawa N., Nonaka M., Wetsel R.A., Colten H.R.;
RT   "Murine complement C2 and factor B genomic and cDNA cloning reveals
RT   different mechanisms for multiple transcripts of C2 and B.";
RL   J. Biol. Chem. 265:19040-19046(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Qin S., Lasky S.R., Loretz C., Dors M., Mahairas G.,
RA   Hood L.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RA   Rowen L., Madan A., Qin S., Hall C., Dahl T., James R., Dickhoff R.,
RA   Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;
RT   "Sequence of the mouse MHC class III region.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 285-761 FROM N.A.
RX   MEDLINE-84061931; PubMed-6689022;
RA   Backstein R., Colten H.R., Woods D.E.;
RT   "Phylogenetic conservation of a class III major histocompatibility
RT   complex antigen, factor B. Isolation and nucleotide sequencing of
RT   mouse factor B cDNA clones.";
RL   J. Biol. Chem. 258:14693-14697(1983).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha-chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha-chain to yield C5a and C5b. Both
CC       cleavages take place at the C-terminal of an arginine residue.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------- ... ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ... ----------
DR   EMBL; M60646; AAA37379.1; -.
DR   EMBL; M60629; AAA37379.1; JOINED.
DR   EMBL; M60630; AAA37379.1; JOINED.
DR   EMBL; M60631; AAA37379.1; JOINED.
DR   EMBL; M60632; AAA37379.1; JOINED.
DR   EMBL; M60633; AAA37379.1; JOINED.
DR   EMBL; M60634; AAA37379.1; JOINED.
DR   EMBL; M60635; AAA37379.1; JOINED.
DR   EMBL; M60636; AAA37379.1; JOINED.
DR   EMBL; M60637; AAA37379.1; JOINED.
DR   EMBL; M60638; AAA37379.1; JOINED.
DR   EMBL; M60639; AAA37379.1; JOINED.
DR   EMBL; M60640; AAA37379.1; JOINED.
DR   EMBL; M60641; AAA37379.1; JOINED.
DR   EMBL; M60642; AAA37379.1; JOINED.
DR   EMBL; M60643; AAA37379.1; JOINED.
DR   EMBL; M60644; AAA37379.1; JOINED.
DR   EMBL; M60645; AAA37379.1; JOINED.
DR   EMBL; AF049850; AAC05283.1; -.
```

```
DR   EMBL; AF109906; AAC84166.1; -.
DR   EMBL; K01406; AAA39649.1; -.
DR   EMBL; K01407; AAA39650.1; -.
DR   EMBL; K01408; AAA39651.1; -.
DR   EMBL; M57890; AAA63263.1; -.
DR   PIR; A38875; BBMS.
DR   HSSP; P00734; 2HNT.
DR   MEROPS; S01.196; -.
DR   MGD; MGI:105975; H2 Bf.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 3.
DR   Pfam; PF00089; Trypsin; 1.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    761       COMPLEMENT FACTOR B.
FT   CHAIN        23    256       BA FRAGMENT.
FT   CHAIN       257    761       BB FRAGMENT.
FT   DOMAIN       33     90       SUSHI 1.
FT   DOMAIN       99    156       SUSHI 2.
FT   DOMAIN      161    216       SUSHI 3.
FT   DOMAIN      267    460       VWFA.
FT   DOMAIN      479    761       SERINE PROTEASE.
FT   ACT_SITE    523    523       CHARGE RELAY SYSTEM.
FT   ACT_SITE    573    573       CHARGE RELAY SYSTEM.
FT   ACT_SITE    696    696       CHARGE RELAY SYSTEM.
FT   DISULFID     34     79       BY SIMILARITY.
FT   DISULFID     59     95       BY SIMILARITY.
FT   DISULFID    100    142       BY SIMILARITY.
FT   DISULFID    128    156       BY SIMILARITY.
FT   DISULFID    162    202       BY SIMILARITY.
FT   DISULFID    188    215       BY SIMILARITY.
FT   CARBOHYD    119    119       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    282    282       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    730    730       L -> P (IN REF. 4).
FT   CONFLICT    745    745       V -> E (IN REF. 4).
SQ   SEQUENCE   761 AA;  85004 MW;  9E9015448A61020E CRC64;

  Query Match          16.2%;  Score 122;  DB 1;  Length 761;
  Best Local Similarity   28.3%;  Pred. No. 9.2e-05;
  Matches   38;  Conservative   16;  Mismatches   64;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNG    RISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWD 57
           | |  |    ||    |  ||:    :   | : |   : | |  :  |    : :| ||
Db      98 IRCPRPQDEFNGEEWPRSPEYN    LSDQISFQCYDGYVLRGSANRTC   QENGRWD 149

Qy      58 KPAPKCEYFNKYSSCPEPIVNGGYKIRGSTPYRHGDSVIFACKTNFSMNGNKSVWCQANN 117
                :    |  || | :| | :  ||  ||  | ||: |     : |::   ||
Db     150 GQTAICDDGAGY  CENVLIPIGIRKVGS QYRLEDIVYHCSRGLVLRGSQKRKCQEGG 206

Qy     118 MWGLTRLPTCVSVF 131
             :  |  |:|  |
Db     207 SWSGIE PSCQDSF 219


Search completed: November 6, 2002, 16:05:27
Job time : 13.4684 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd

SM protein protein search, using sw model

Run on: November 6, 2002, 16:05:00 ; Search time 21.4201 Seconds
(without alignments)
1082.224 Million cell updates/sec

Title: US-09-834-309-4
Perfect score: 751
Sequence: 1 GISCGSPPPILNGRISYYST.....ANNMWGPTRLPTCVSVFPLE 134

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	751	100.0	1042	4	Q13866	Q13866 homo sapien
2	751	100.0	1087	4	Q14212	Q14212 homo sapien
3	466.5	62.1	679	11	Q99254	Q99254 mus musculu
4	466.5	62.1	1042	11	Q9DC83	Q9dc83 mus musculu
5	464.5	61.9	1045	6	O46545	O46545 ovis aries
6	235	31.3	1911	6	Q29528	Q29528 papio hamad
7	220	29.3	2039	4	Q16745	Q16745 homo sapien
8	220	29.3	2489	4	Q16744	Q16744 homo sapien
9	218	29.0	142	4	Q9HB99	Q9hb99 homo sapien
10	218	29.0	142	4	Q9HB98	Q9hb98 homo sapien
11	218	29.0	2014	6	Q29530	Q29530 pan troglod
12	215	28.6	132	4	Q9HC55	Q9hc55 homo sapien
13	215	28.6	259	12	P87616	P87616 cowpox viru
14	211	28.1	142	4	Q9HC54	Q9hc54 homo sapien
15	210	28.0	263	12	Q89859	Q89859 variola vir
16	210	28.0	264	12	Q07033	Q07033 variola vir
17	208	27.7	263	12	Q89076	Q89076 variola vir
18	208	27.7	263	12	Q89061	Q89061 variola vir
19	203.5	27.1	395	12	Q9J2M6	Q9j2m6 macaca mula
20	203.5	27.1	579	11	Q60736	Q60736 mus musculu
21	203	27.0	522	6	Q28769	Q28769 papio cynoc
22	199	26.5	559	4	Q9UQV2	Q9uqv2 homo sapien
23	198.5	26.4	533	11	O08569	O08569 cavia porce
24	198	26.4	522	6	Q28797	Q28797 pan troglod
25	196	26.1	661	6	Q29531	Q29531 pan troglod
26	194	25.8	315	6	Q28770	Q28770 papio cynoc
27	194	25.8	645	12	Q9WR72	Q9wr72 macaca mula
28	191.5	25.5	555	11	Q99JA1	Q99ja1 cavia porce
29	190	25.3	363	6	O02839	O02839 sus scrofa
30	188.5	25.1	451	13	Q9DEG0	Q9deg0 gallus gall
31	184	24.5	417	11	O35520	O35520 rattus norv
32	184	24.5	497	11	Q63612	Q63612 rattus norv
33	184	24.5	559	11	Q63135	Q63135 rattus norv
34	180.5	24.0	483	11	Q64735	Q64735 mus musculu
35	180.5	24.0	657	4	Q14006	Q14006 homo sapien
36	178.5	23.8	469	11	Q91X48	Q91x48 mus musculu
37	176	23.4	337	11	P97263	P97263 cavia porce
38	175.5	23.4	1172	4	Q9NU87	Q9nu87 homo sapien
39	172	22.9	300	11	Q9R0R9	Q9r0r9 mus musculu
40	171	22.8	267	11	P97939	P97939 cavia porce
41	171	22.8	339	11	P97261	P97261 cavia porce
42	171	22.8	379	11	P70105	P70105 cavia porce
43	169.5	22.6	268	12	Q9QAX6	Q9qax6 yaba monkey
44	169.5	22.6	550	12	P88903	P88903 kaposi's sa
45	168.5	22.4	550	12	O40912	O40912 kaposi's sa
46	168	22.4	365	11	O88174	O88174 mus musculu
47	168	22.4	481	4	Q9H284	Q9h284 homo sapien
48	167.5	22.3	390	11	Q921P0	Q921p0 mus musculu
49	167.5	22.3	699	11	Q91WX0	Q91wx0 rattus norv
50	167.5	22.3	1236	11	Q91YB6	Q91yb6 rattus norv
51	167.5	22.3	3567	11	Q9ES77	Q9es77 mus musculu
52	166.5	22.2	569	4	Q9BXR6	Q9bxr6 homo sapien
53	165.5	22.0	808	11	Q61408	Q61408 mus musculu
54	165	22.0	225	12	Q91MM3	Q91mm3 lumpy skin
55	164	21.8	2043	5	O96943	O96943 geodia cydo
56	163	21.7	285	6	O19121	O19121 papio hamad
57	162	21.6	343	11	Q61406	Q61406 mus musculu
58	161.5	21.5	305	6	Q9MYJ5	Q9myj5 pan troglod
59	161.5	21.5	347	6	Q9MYJ6	Q9myj6 pan troglod
60	160.5	21.4	347	6	Q9MYJ7	Q9myj7 gorilla gor
61	160	21.3	285	6	O19126	O19126 macaca fasc
62	160	21.3	285	6	O19127	O19127 macaca mula
63	160	21.3	369	6	P79138	P79138 cercopithec
64	160	21.3	754	13	Q98U16	Q98u16 cyprinus ca
65	159	21.2	3389	4	Q96QU9	Q96qu9 homo sapien
66	159	21.2	3508	4	Q96RM4	Q96rm4 homo sapien
67	158	21.0	336	6	O62834	O62834 saguinus oe
68	157	20.9	372	11	Q9QYJ9	Q9qyj9 rattus norv
69	157	20.9	378	6	O62837	O62837 saguinus oe
70	157	20.9	382	11	Q9QUP3	Q9qup3 rattus norv
71	157	20.9	399	11	Q9QUN6	Q9qun6 rattus norv
72	157	20.9	399	11	Q9Z0M0	Q9z0m0 rattus norv
73	157	20.9	466	11	Q9Z0L9	Q9z0l9 rattus norv
74	156	20.8	1001	4	Q9Y2E1	Q9y2e1 homo sapien
75	155	20.6	349	4	Q15429	Q15429 homo sapien
76	154.5	20.6	3564	11	Q923L3	Q923l3 mus musculu
77	154	20.5	222	6	O19128	O19128 pithecia pi
78	153	20.4	222	6	O19124	O19124 saguinus my
79	153	20.4	314	6	O62835	O62835 saguinus oe
80	153	20.4	377	6	O62838	O62838 saguinus oe
81	153	20.4	737	13	Q90422	Q90422 brachydanio
82	153	20.4	1061	5	Q9VXX7	Q9vxx7 drosophila
83	152.5	20.3	343	6	Q9MYJ4	Q9myj4 papio hamad
84	152.5	20.3	946	4	Q96QQ3	Q96qq3 homo sapien
85	152	20.2	762	13	Q9YIC6	Q9yic6 cyprinus ca
86	151	20.1	1653	5	Q9VIU9	Q9viu9 drosophila
87	149	19.8	809	4	Q9BW82	Q9bw82 homo sapien
88	149	19.8	840	4	Q9UJ45	Q9uj45 homo sapien
89	149	19.8	853	4	Q9UJ47	Q9uj47 homo sapien

90	148.5	19.8	344	5	Q9MYJ3	Q9myj3 macaca mula
91	148	19.7	772	4	Q9NUI3	Q9nui3 homo sapien
92	148	19.7	787	4	Q9NUI4	Q9nui4 homo sapien
93	148	19.7	792	4	Q9Y336	Q9y3j6 homo sapien
94	148	19.7	923	11	Q91X64	Q91x64 mus musculu
95	148	19.7	1024	1	Q9BYH1	Q9byh1 homo sapien
96	147	19.6	267	12	Q9DHG9	Q9dhg9 yaba like d
97	146	19.4	222	5	Q19122	Q19122 callimico g
98	144.5	19.2	271	5	Q9MYJ0	Q9myj0 erythrocebu
99	144.5	19.2	278	5	Q9MYJ1	Q9myj1 erythrocebu
100	144.5	19.2	344	5	Q9MYJ2	Q9myj2 erythrocebu

ALIGNMENTS

RESULT 1
Q13866
ID Q13866 PRELIMINARY; PRT; 1032 AA.
AC Q13866;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CR2 RECEPTOR PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B LYMPHOCYTES OF 3 DONORS;
RX MEDLINE=88171282; PubMed=2832506;
RA Weis J.J., Toothaker L.E., Smith J.A., Weis J.H., Fearon D.T.;
RT "Structure of the human B lymphocyte receptor for C3d and the Epstein-
RT Barr virus and relatedness to other members of the family of C3/C4
RT binding proteins.";
RL J. Exp. Med. 167:1047-1066(1988).
DR EMBL; Y00649; CAA68674.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000884; Zn_carbopept.
DR Pfam; PF00084; sushi; 15.
DR SMART; SM00032; CCP; 14.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW Signal; Receptor.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1032 POTENTIAL.
SQ SEQUENCE 1032 AA; 112844 MW; 14F7D82CA6A1B888 CRC64;

Query Match 100.0%; Score 751; DB 4; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.5e-74;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GISCGSPPPILNGRISYYSIPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 GISCGSPPPILNGRISYYSIPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 79

Qy      61 PKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 PKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 139

Qy     121 PTRLPTCVSVFPLE 134
           ||||||||||||||
Db     140 PTRLPTCVSVFPLE 153
```

RESULT 2
Q14212
ID Q14212 PRELIMINARY; PRT; 1087 AA.
AC Q14212;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CR2 PROTEIN PRECURSOR.
GN CR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88097454; PubMed=2827171;
RA Moore M., Cooper N., Tack B., Nemerow G.;
RT "Molecular cloning of the cDNA encoding the epstein barr virus/C3d
RT receptor (complement receptor type 2) of human b lymphocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198(1987).
DR EMBL; J03565; AAA35784.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000884; Zn_carbopept.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 14.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1087 POTENTIAL.
SQ SEQUENCE 1087 AA; 118933 MW; CB2865350291788 CRC64;

Query Match 100.0%; Score 751; DB 4; Length 1087;
Best Local Similarity 100.0%; Pred. No. 1.6e-74;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GISCGSPPPILNGRISYYSIPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 GISCGSPPPILNGRISYYSIPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 79

Qy      61 PKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 PKCEYFNKYSSCPEPIVPGGYKIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 139

Qy     121 PTRLPTCVSVFPLE 134
           ||||||||||||||
Db     140 PTRLPTCVSVFPLE 153
```

RESULT 3
Q99254
ID Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christenson S.M., Weis J.H.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1.";
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.U., Kinoshita T., Molina H., Bourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein.";
RL J. Exp. Med. 181:151-159(1995).

```
DR   EMBL; U17128; AAA78271.1; -.
DR   EMBL; U17123; AAA78271.1; JOINED.
DR   EMBL; U17124; AAA78271.1; JOINED.
DR   EMBL; U17125; AAA78271.1; JOINED.
DR   EMBL; U17126; AAA78271.1; JOINED.
DR   EMBL; U17127; AAA78271.1; JOINED.
DR   EMBL; M86470; AAA37449.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88489; Cr2.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 10.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 10.
FT   NON_TER     679     679
SQ   SEQUENCE   679 AA;  74916 MW;  52FC00FDCED20CDC CRC64;

  Query Match             62.1%;  Score 466.5;  DB 11;  Length 679;
  Best Local Similarity   61.2%;  Pred. No. 3.2e-43;
  Matches   82;  Conservative   18;  Mismatches   33;  Indels    1;  Gaps    1;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           ||| || : | |  ||| || |||:||:|| ::|||||:: ||:::|  |||| |
Db     397 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 456

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
             || ||  :| :|||||:  :|| |:|||||||| || ||:| |:|:|||||| |||
Db     457 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 516

Qy     121 PTRLPTCVSVFPLE 134
           || || | | ||||
Db     517 PTALPVCESDFPLE 530


RESULT 4
Q9DC83
ID   Q9DC83      PRELIMINARY;      PRT;  1032 AA.
AC   Q9DC83;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 2.
GN   CR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SPLEEN;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriquez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004072; BAB22548.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88489; Cr2.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 14.
DR   SMART; SM00032; CCP; 14.
SQ   SEQUENCE   1032 AA;  113335 MW;  6F96C3D95E431E89 CRC64;

  Query Match             62.1%;  Score 466.5;  DB 11;  Length 1032;
  Best Local Similarity   61.2%;  Pred. No. 5.3e-43;
  Matches   82;  Conservative   18;  Mismatches   33;  Indels    1;  Gaps    1;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           ||| || : | |  ||| || |||:||:|| ::|||||:: ||:::|  |||| |
Db      21 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 80

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
             || ||  :| :|||||:  :|| |:|||||||| || ||:| |:|:|||||| |||
Db      81 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 140

Qy     121 PTRLPTCVSVFPLE 134
           || || | | ||||
Db     141 PTALPVCESDFPLE 154


RESULT 5
O46545
ID   O46545      PRELIMINARY;      PRT;  1045 AA.
AC   O46545;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 2.
GN   CR2.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE ALPINE;
RX   MEDLINE=98309471; PubMed=9647256;
RA   Hein W.R., Dudler L., Marston W.L., Tardsverk T., Young A.J.,
RA   Avila D.;
RT   "Ubiquitination and dimerization of complement receptor type 2 on
RT   sheep B cells.";
RL   J. Immunol. 161:458-466(1998).
DR   EMBL; AF038131; AAB92375.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 14.
DR   SMART; SM00032; CCP; 14.
KW   Receptor.
SQ   SEQUENCE   1045 AA;  115562 MW;  FF58E A2892C0D59 CRC64;

  Query Match             61.9%;  Score 464.5;  DB 6;  Length 1045;
  Best Local Similarity   59.3%;  Pred. No. 9e-43;
  Matches   80;  Conservative   18;  Mismatches   36;  Indels    1;  Gaps    1;

Qy       1 GISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 60
           || |  || | |||  |:| |:|:  |: |||   |||||:| | ||:|| |:| |||
Db      29 GIFCDPPPESIKNGRSGYHSGPVAINAVVTYSCASAFRLIGERRLFCISKDNVNGIWDKAP 88

Qy      61 PKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 119
           | |||:| | | |||:|||| : |  :|||:|||:  ||:| |||:|||||| |
Db      89 PICEYYNRNSVCSEPIIPGGYRTKTSRPPIRHGDAVTFSCNANFTMKGNKTVWCQANRKW 148

Qy     120 GPTRLPTCVSVFPLE 134
           ||| || | |  |||
Db     149 GPTPLPICESDTPLE 163


RESULT 6
Q29528
```

```
ID   Q29528       PRELIMINARY;      PRT;  1911 AA.
AC   Q29528;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT COMPONENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.;
RA   Atkinson J.P.;
RT   "Primary sequence of the baboon 200 kDa C4b/C4b receptor (CR1).";
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L39791; AAA62170.1; -.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 29.
DR   SMART; SM00032; CCP; 29.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_4.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
FT   NON_TER    1911   1911
SQ   SEQUENCE   1911 AA;  216174 MW;  535A4DD0EAFA521D CRC64;

  Query Match             31.8%;  Score 235;  DB 6;  Length 1911;
  Best Local Similarity   38.8%;  Pred. No. 5.2e-17;
  Matches   54;  Conservative   16;  Mismatches   53;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC       SGTFRLIGEKSLICTKDKV 52
           ||| :|: | ||   :||   |    |||: | |          | |:|||:|: | :||
Db    1404 ISCKPPPTISNG--DFYSNNKTSFHSGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1461

Qy      53 DGTWDEPAPKCEYPNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKTNESMNGNKSV 111
           | |  | |:|   ||   | | | | :: |:   :   : | | |:  | | |: :|
Db    1462 VGAWSSPPPRCISINK--CTAPEVKNGIRVPGNRSFFSLNEIVRFRCQPGFVMVGSHTV 1518

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || || ||| :|| | |
Db    1519 QCQTNNRWGP-KLPHCSRV 1536


RESULT 7
Q16745
ID   Q16745       PRELIMINARY;      PRT;  2039 AA.
AC   Q16745;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94065175; PubMed=8245463;
RA   Vik D.P., Wong W.W.;
RT   "Structure of the gene for the F allele of complement receptor type 1
RT   and sequence of the coding region unique to the S allele.";
RL   J. Immunol. 151:6214-6224(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Vik D.P., Wong W.W.;
RL   Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L17418; AAB60694.1; -.
DR   EMBL; L17390; AAB60694.1; JOINED.
DR   EMBL; L17399; AAB60694.1; JOINED.
DR   EMBL; L17409; AAB60694.1; JOINED.
DR   EMBL; L17419; AAB60694.1; JOINED.
DR   EMBL; L17420; AAB60694.1; JOINED.
DR   EMBL; L17421; AAB60694.1; JOINED.
DR   EMBL; L17422; AAB60694.1; JOINED.
DR   EMBL; L17423; AAB60694.1; JOINED.
DR   EMBL; L17341; AAB60694.1; JOINED.
DR   EMBL; L17392; AAB60694.1; JOINED.
DR   EMBL; L17393; AAB60694.1; JOINED.
DR   EMBL; L17394; AAB60694.1; JOINED.
DR   EMBL; L17395; AAB60694.1; JOINED.
DR   EMBL; L17396; AAB60694.1; JOINED.
DR   EMBL; L17397; AAB60694.1; JOINED.
DR   EMBL; L17398; AAB60694.1; JOINED.
DR   EMBL; L17400; AAB60694.1; JOINED.
DR   EMBL; L17401; AAB60694.1; JOINED.
DR   EMBL; L17402; AAB60694.1; JOINED.
DR   EMBL; L17403; AAB60694.1; JOINED.
DR   EMBL; L17404; AAB60694.1; JOINED.
DR   EMBL; L17405; AAB60694.1; JOINED.
DR   EMBL; L17406; AAB60694.1; JOINED.
DR   EMBL; L17407; AAB60694.1; JOINED.
DR   EMBL; L17408; AAB60694.1; JOINED.
DR   EMBL; L17410; AAB60694.1; JOINED.
DR   EMBL; L17411; AAB60694.1; JOINED.
DR   EMBL; L17412; AAB60694.1; JOINED.
DR   EMBL; L17413; AAB60694.1; JOINED.
DR   EMBL; L17414; AAB60694.1; JOINED.
DR   EMBL; L17415; AAB60694.1; JOINED.
DR   EMBL; L17416; AAB60694.1; JOINED.
DR   EMBL; L17417; AAB60694.1; JOINED.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   2039 AA;  223603 MW;  B82FCB11C68B1645 CRC64;

  Query Match             29.4%;  Score 220;  DB 4;  Length 2039;
  Best Local Similarity   37.4%;  Pred. No. 2.6e-15;
  Matches   52;  Conservative   16;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYS---TPIAVGTVIRYSC       SGTFRLIGEKSLICTKDKV 52
           ||| :|: | ||   :||   |    |||: | |          | |:|||:|: | :||
Db    1517 ISCEPPPTISNG--DFYSNNKTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 DGTWDEPAPKCEYPNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKTNESMNGNKSV 111
           | |  | |:|   ||   | | | | :: |:   :   : | | |:  | | |: :|
Db    1575 VGVWSSPPPRCISINK--CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 1631

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| | |
Db    1632 QCQTNGRWGP-KLPHCSRV 1649


RESULT 8
Q16744
ID   Q16744       PRELIMINARY;      PRT;  2489 AA.
AC   Q16744;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94065175; PubMed=8245463;
RA   Vik D.P., Wong W.W.;
RT   "Structure of the gene for the F allele of complement receptor type 1
RT   and sequence of the coding region unique to the S allele.";
RL   J. Immunol. 151:6214-6224(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Vik D.P., Wong W.W.;
RL   Submitted (JUN-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L17418; AAB60695.1; -.
DR   EMBL; L17390; AAB60695.1; JOINED.
DR   EMBL; L17399; AAB60695.1; JOINED.
DR   EMBL; L17409; AAB60695.1; JOINED.
DR   EMBL; L17419; AAB60695.1; JOINED.
DR   EMBL; L17420; AAB60695.1; JOINED.
DR   EMBL; L17421; AAB60695.1; JOINED.
DR   EMBL; L17422; AAB60695.1; JOINED.
DR   EMBL; L17423; AAB60695.1; JOINED.
DR   EMBL; L17424; AAB60695.1; JOINED.
DR   EMBL; L17425; AAB60695.1; JOINED.
DR   EMBL; L17426; AAB60695.1; JOINED.
DR   EMBL; L17427; AAB60695.1; JOINED.
DR   EMBL; L17428; AAB60695.1; JOINED.
DR   EMBL; L17429; AAB60695.1; JOINED.
DR   EMBL; L17430; AAB60695.1; JOINED.
DR   EMBL; L17391; AAB60695.1; JOINED.
DR   EMBL; L17392; AAB60695.1; JOINED.
DR   EMBL; L17393; AAB60695.1; JOINED.
DR   EMBL; L17394; AAB60695.1; JOINED.
DR   EMBL; L17395; AAB60695.1; JOINED.
DR   EMBL; L17396; AAB60695.1; JOINED.
DR   EMBL; L17397; AAB60695.1; JOINED.
DR   EMBL; L17398; AAB60695.1; JOINED.
DR   EMBL; L17400; AAB60695.1; JOINED
DR   EMBL; L17401; AAB60695.1; JOINED.
DR   EMBL; L17402; AAB60695.1; JOINED.
DR   EMBL; L17403; AAB60695.1; JOINED.
DR   EMBL; L17404; AAB60695.1; JOINED.
DR   EMBL; L17405; AAB60695.1; JOINED.
DR   EMBL; L17406; AAB60695.1; JOINED.
DR   EMBL; L17407; AAB60695.1; JOINED.
DR   EMBL; L17408; AAB60695.1; JOINED.
DR   EMBL; L17410; AAB60695.1; JOINED.
DR   EMBL; L17411; AAB60695.1; JOINED.
DR   EMBL; L17412; AAB60695.1; JOINED.
DR   EMBL; L17413; AAB60695.1; JOINED.
DR   EMBL; L17414; AAB60695.1; JOINED.
DR   EMBL; L17415; AAB60695.1; JOINED.
DR   EMBL; L17416; AAB60695.1; JOINED.
DR   EMBL; L17417; AAB60695.1; JOINED.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 37.
DR   SMART; SM00032; CCP; 37.
DR   PROSITE; PS00143; CARBOXYPEPT_ZN_2; UNKNOWN_3.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   2489 AA;  272846 MW;  CEE11B53F2B4FAF6 CRC64;

  Query Match             29.4%;  Score 220;  DB 4;  Length 2489;
  Best Local Similarity   37.4%;  Pred. No. 3.3e-15;
  Matches   52;  Conservative   16;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |    |||: | |         | |:||:|: | :||
Db    1967 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 2024

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
             | | | |:|   ||   | | |    :: |:  :   : | | |:  | | |: :|
Db    2025 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTFIVRFRCQPGFVMVGSHTV 2081

Qy     112 WCQANNMWGFTRLPTCVSV 130
           || |   || :|| | |
Db    2082 QCQTNGRWGF-KLPHCSRV 2099


RESULT 9
Q9HB99
ID   Q9HB99       PRELIMINARY;      PRT;   132 AA.
AC   Q9HB99;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Miller L.H., Doumbo O., Moulds J.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264715; AAG14442.1; -.
DR   HSSP; P02749; 1QUB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1       1
FT   VARIANT      94      94       T -> S.
FT   NON_TER     132     132
SQ   SEQUENCE   132 AA;  14680 MW;  71A9B91B5596958B CRC64;

  Query Match             29.0%;  Score 218;  DB 4;  Length 132;
  Best Local Similarity   36.8%;  Pred. No. 1.6e-15;
  Matches   50;  Conservative   17,  Mismatches   53;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |    |||: | |         | |:||:|: | :||
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSV 111
             | | | |:|   ||   | | |    :: |:  :   : | | |:  | | |: :|
Db      59 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFTLIEIIRFRCQPGFVMVGSHTV 115

Qy     112 WCQANNMWGFTRLPTC 127
           || |   || :|| |
Db     116 QCQTNGRWGF-KLPHC 130


RESULT 10
Q9HB98
ID   Q9HB98       PRELIMINARY;      PRT;   132 AA.
AC   Q9HB98;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Miller L.H., Doumbo O., Moulds J.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264716; AAG14443.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
```

```
KW   Receptor.
FT   NON_TER       1       1
FT   NON_TER     142     142
SQ   SEQUENCE   142 AA;  14666 MW;  1AC2B91B659691DA CRC64;

  Query Match             29.0%;  Score 218;  DB 4;  Length 142;
  Best Local Similarity   36.8%;  Pred. No. 1.6e-16;
  Matches   50;  Conservative   17;  Mismatches   53;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPFILNGRISYYS  TPIAVGTVIRYSC  SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |   ||| | |        | |:|||:| | :||
Db       1 ISCEPPPFIISNG  DFYSNNRISFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 DGIWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKINFSMNGNKSV 111
           | | | |:|   ||   | | |   :: |:   :   : : | |:  | | |: :|
Db      59 VGVWSSPPPRCISTNK  CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 115

Qy     112 WCQANNMWGPTRLPTC 127
           | | |  ||| :|| |
Db     116 QCQTNGRWGP KLPHC 130


RESULT 11
Q29530
ID   Q29530       PRELIMINARY;       PRT;  2014 AA.
AC   Q29530;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262799; PubMed=8021505;
RA   Birmingham D.J., Shen X.L., Boutcade D., Nickells M.W., Atkinson J.P.;
RT   "Primary sequence of an alternatively spliced form of CR1. Candidate
RT   for the 75,000 M(r) complement receptor expressed on chimpanzee
RT   erythrocytes.";
RL   J. Immunol. 153:691-700(1994).
DR   EMBL; L24920; AAA51438.1; -.
DR   HSSP; P08603; 1HFL.
DR   InterPro; IPR001024; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000884; Zn_carbopept.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
DR   PROSITE; PS00143; CARBOXYPEPT_ZN_2; UNKNOWN 2.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN 1.
KW   Receptor.
FT   NON_TER       1       1
FT   NON_TER    2014    2014
SQ   SEQUENCE   2014 AA;  221781 MW;  6D0C3A74DB1F1D89 CRC64;

  Query Match             29.0%;  Score 218;  DB 6;  Length 2014;
  Best Local Similarity   37.4%;  Pred. No. 4.3e-15;
  Matches   52;  Conservative   16;  Mismatches   55;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPFILNGRISYYSITPIAV  GTVIRYSC  SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |   ||| | |        | |:|||:| | :||
Db    1492 ISCEPPPFIISNG  DFYSNNRASFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1549

Qy      53 DGIWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKINFSMNGNKSV 111
           | | | |:|   ||   | | |   :: |:   :   : : | |:  | | |: :|
Db    1550 VGVWSSPPPRCISTNK  CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHIV 1606

Qy     112 WCQANNMWGPTRLPTCVSV 130
           || |  ||| :|| |  |
Db    1607 QCQTNGRWGP KLPHCSRV 1624


RESULT 12
Q9HC55
ID   Q9HC55       PRELIMINARY;       PRT;  132 AA.
AC   Q9HC55;
DT   01-MAR-2000 (TrEMBLrel. 16, Created)
DT   01-MAR-2000 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Birmingham D., Atkinson J.P.;
RA   Miller L., Doumbo O., Moulds J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF169969; AAG09139.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1       1
FT   VARIANT      85      85       G -> R.
FT   NON_TER     132     132
SQ   SEQUENCE   132 AA;  14553 MW;  1AD11B90F59691CB CRC64;

  Query Match             28.6%;  Score 215;  DB 4;  Length 132;
  Best Local Similarity   37.5%;  Pred. No. 3.4e-16;
  Matches   51;  Conservative   15;  Mismatches   54;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPFILNGRISYYS---TPIAVGTVIRYSC-      SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |   ||| | |        | |:|||:| | :||
Db       1 ISCEPPPFIISNG  DFYSNNRISFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 DGIWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKINFSMNGNKSV 111
           | | | |:|   ||   | | |   :: |:   :   : : | |:  | | |: :|
Db      59 VGVWSSPPPRCISTNK---CTAPEVENAIGVPGNRSFFSLTEIVRFRCQPGFVMVGSHIV 115

Qy     112 WCQANNMWGPTRLPTC 127
           || |  ||| :|| |
Db     116 QCQTNGRWGP KLPHC 130


RESULT 13
P87616
ID   P87616       PRELIMINARY;       PRT;  259 AA.
AC   P87616;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   41KBP FRAGMENT FROM LEFT END OF GENOME.
GN   D17L OR C17L.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=97068532; PubMed=8963248;
RA   Safronov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,
RA   Shchelkunov S.N., Sandakhchiev L.S.;
RT   "Genes of a circle of hosts for the cowpox virus.";
RL   Dokl. Akad. Nauk 349:829-833(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=98229462; PubMed=9568042;
RA   Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A.,
```

```
RA   Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
RT   "The genomic sequence analysis of the left and right species-specific
RT   terminal region of a cowpox virus strain reveals unique sequences and
RT   a cluster of intact ORFs for immunomodulatory and host range
RT   proteins.";
RL   Virology 243:432-460(1998).
DR   EMBL; X94355; CAA64102.1; -.
DR   EMBL; Y11842; CAA72567.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   259 AA;  28193 MW;  9D1AAEF6893B859A CRC64;

  Query Match             28.6%;  Score 215;  DB 12;  Length 259;
  Best Local Similarity   35.9%;  Pred. No. 7.7e-16;
  Matches   46;  Conservative   15;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGKISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGIWDKPAP 61
           : | ||| | ||| : |      |:|: |||:  : |||   :||       | |  | |
Db     142 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLC----SCGEWSDP-P 195

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
            |:       || | :  ||   |    | : |:| | ||  : ::|: |  |   | |
Db     196 TCQIV---KCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWQ 251

Qy     121 PTRLPTCV 128
           |  || ||
Db     252 P-ELPKCV 258


RESULT 14
Q9HC54
ID   Q9HC54      PRELIMINARY;      PRT;   132 AA.
AC   Q9HC54;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Birmingham D., Atkinson J.P.,
RA   Miller L., Doumbo O., Moulds J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF169970; AAG09140.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1      1
FT   VARIANT      74     74       E -> K.
FT   VARIANT      85     85       G -> R.
FT   NON_TER     132    132
SQ   SEQUENCE   132 AA;  14554 MW;  1AD1159EF5989FCB CRC64;

  Query Match             28.1%;  Score 211;  DB 4;  Length 132;
  Best Local Similarity   36.8%;  Pred. No. 9.5e-16;
  Matches   50;  Conservative   16;  Mismatches   54;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPILNGPISYYS---TPIAVGTVIRYSC------SGTFRLIGEKSLLCITKDKV 52
           |||  || | ||   :||   |    |||: | |          | |:|||:| | :||
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVIFACKTNFSMNGNKSV 111
           | |  | |:|   |:   |  | |     : |:   :   : | | |:  | | |: :|
Db      59 VGVWSSPPPRCISINF--QCIAPEVENAIGVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 115

Qy     112 WCQANNMWGPTRLPTC 127
           || +  ||| :|| |
Db     116 QCQTNGRWGP-KLPHC 130


RESULT 15
Q89859
ID   Q89859      PRELIMINARY;      PRT;   263 AA.
AC   Q89859;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HOMOLOG OF VACCINIA VIRUS CDS C3L
GN   D15L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BANGLADESH-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L.I., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SOMALIA-1977;
RA   Massung R.F., Loparev V.N., Knight J. C., Chizhikov V.E., Parsons J.M.,
RA   Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L22579; AAA60760.1; -.
DR   EMBL; U18340; AAA69423.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28816 MW;  04659F8002F74220 CRC64;

  Query Match             28.0%;  Score 210;  DB 12;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 2.8e-15;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGIWDKPAP 61
           : |  || | ||| : |:      |:|: || :  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVIYSCNSGYSLIGNSGVLC----SGGEWSNP-P 199

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
            |:       || | :  ||   |    | : |:| | ||  : ::|: |  |   | |
Db     200 TCQIV---KCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 16
Q07033
ID   Q07033      PRELIMINARY;      PRT;   263 AA.
AC   Q07033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   D12L PROTEIN.
GN   D12L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN INDIA 1967, SSP. MAJOR.
RX   MEDLINE 94202281; PubMed 8304129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA 1967, SSP. MAJOR;
RX   MEDLINE 95159666; PubMed=7856412;
RA   Shchelkunov S.N., Blinov V.M., Resenchuk S.M., Totmenin A.V.,
RA   Olenina L.V., Chirikova O.B., Sandakhchiev L.S.;
RT   "Analysis of the nucleotide sequence of 53 kbp from the right terminus
RT   of the genome of variola major virus strain India 1967.";
RL   Virus Res. 34:207-246(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN INDIA 1967, SSP. MAJOR;
RX   MEDLINE 96320969; PubMed 8752802;
RA   Shchelkunov S.N., Totmenin A.V.;
RT   "Two types of deletions in orthopoxvirus genomes.";
RL   Virus Genes 9:241-245(1995).
DR   EMBL; X69198; CAA48953.1; -.
DR   PIR; C36838; C36838.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28789 MW;  5FE244F48B19D479 CRC64;

  Query Match             28.0%;  Score 210;  DB 12;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 2.8e-15;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 61
           |  || | ||| : :     |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVIC --  SGGEWSNP-P 199

Qy      62 KCEYENEYSSCFEPIVLGYKIEG-SIPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG 120
           |:       || | :  || |   | : |:| | ||  : :::| |  |   | |
Db     200 TCQIV-- KCPHPTILNGYLSSGFKRSYSYNDNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 17
Q89076
ID   Q89076      PRELIMINARY;      PRT;   263 AA.
AC   Q89076;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   B19L PROTEIN.
GN   B19L.
OS   Variola virus, and
OS   variola minor virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255, 54258;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Variola virus; STRAIN=GARCIA 1966;
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=variola minor virus; STRAIN=GARCIA-1966;
RA   Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Esposito J.J., Sosnovtsev S.;
RT   "Analysis of the complete coding sequence of DNA of alastrim variola
RT   minor virus strain Garcia-1966.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U18338; AAA69458.1; -.
DR   EMBL; Y16780; CAB54612.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28842 MW;  E5E51D91027C423D CRC64;

  Query Match             27.7%;  Score 208;  DB 12;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 4.7e-15;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 61
           : :  || | ||| : :     |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVIC --  SGGEWSNP-P 199

Qy      62 KCEYENEYSSCFEPIVLGYKIEG-SIPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG 120
           |:       || | :  || |   | : |:| | ||  : :::| |  |   | |
Db     200 TCQIV-- KCPHPTILNGYLSSGFKRSYSYNDNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262


RESULT 18
Q89061
ID   Q89061      PRELIMINARY;      PRT;   263 AA.
AC   Q89061;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D15L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CONGO-1965;
RX   MEDLINE=96295428; PubMed=8661439;
RA   Massung R.F., Loparev V.N., Knight J.C., Totmenin A.V.,
RA   Chizhikov V.E., Parsons J.M., Safronov P.F., Gutorov V.V.,
RA   Shchelkunov S.N., Esposito J.J.;
RT   "Terminal region sequence variations in variola virus DNA.";
RL   Virology 221:291-300(1996).
DR   EMBL; U18337; AAA69317.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28843 MW;  04650C966EAC4220 CRC64;

  Query Match             27.7%;  Score 208;  DB 12;  Length 263;
  Best Local Similarity   35.2%;  Pred. No. 4.7e-15;
  Matches   45;  Conservative   16;  Mismatches   55;  Indels   12;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 61
           : :  || | ||| : :     |:|: |||:  : |||   :||       | |  | |
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVIC --  SGGEWSNP-P 199

Qy      62 KCEYENEYSSCFEPIVLGYKIEG-SIPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG 120
           |:       || | :  || |   | : |:| | ||  : :::| |  |   | |
Db     200 TCQIV-- KCPHPTILNGYLSSGFKRSYSYNDNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     121 PTRLPTCV 128
           |  || ||
Db     256 P-ELPKCV 262
```

```
RESULT 19
Q9QJ2M6
ID   Q9QJ2M6      PRELIMINARY;      PRT;   395 AA.
AC   Q9QJ2M6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT BINDING PROTEIN.
OS   Macaca mulatta rhadinovirus 26-95.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=119193;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MACACA MULATTA RHADINOVIRUS ISOLATE 26-95;
RX   MEDLINE=20173730; PubMed=10708456;
RA   Alexander L., Denekamp L., Knapp A., Auerbach M.R., Damania B.,
RA   Desrosiers R.C.;
RT   "The primary sequence of rhesus monkey rhadinovirus isolate 26-95:
RT   sequence similarities to Kaposi's sarcoma-associated herpesvirus and
RT   rhesus monkey rhadinovirus isolate 17577.";
RL   J. Virol. 74:3388-3398(2000).
DR   EMBL; AF210726; AAF59982.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR001230; Prenyltn.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ   SEQUENCE   395 AA;  43922 MW;  B4C9C6F2E226AE06 CRC64;

  Query Match             27.1%;  Score 203.5;  DB 12;  Length 395;
  Best Local Similarity   34.1%;  Pred. No. 2.4e-14;
  Matches   46;  Conservative   22;  Mismatches   42;  Indels   25;  Gaps    7;

Qy       4 PGSPPPILNGR-------ISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTW 56
           |   | | ||:        :|  |       | ::|:  | |||  :  |:|    :|||
Db     156 CKKQPDIENGKYYPVQEFYNYLET------ITFTCNKDFSLIGNTTTTCMT----NGTW 204

Qy      57 DKPAPKCEYFNKYSSCPEPIVPGGYKIRGST-PYRHGDSVTFACKTNFSMNGNKSVWCQA 115
             | |||:      :|  | :  |  : ||:  |:|| |||  |:  |::||:|   |:
Db     205 SSPVPKCQQI----TCSAPNIDHGTLIVGSSPVYKHGQSVTIGCEDGFTLNGHKMCTCE- 259

Qy     116 NNMWGPTRLPTCVSV 130
             ::| |  ||||| :
Db     260 YSLWNPP-LPTCVPI 273


RESULT 20
Q60736
ID   Q60736      PRELIMINARY;      PRT;   579 AA.
AC   Q60736;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPERM FERTILIZATION PROTEIN SP56 PRECURSOR.
GN   ZP4R OR SP56.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HSD/ICR; TISSUE=TESTIS;
RX   MEDLINE=95427956; PubMed=7604284;
RA   Bookbinder L.H., Cheng A., Bleil J.D.;
RT   "Tissue- and species-specific expression of sp56, a mouse sperm
RT   fertilization protein.";
RL   Science 269:86-89(1995).
DR   EMBL; U17108; AAC52208.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:104965; Zp3r.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
KW   Signal.
FT   SIGNAL        1     32       POTENTIAL.
FT   CHAIN        33    579       SPERM FERTILIZATION PROTEIN SP56.
SQ   SEQUENCE   579 AA;  64950 MW;  C87BD1DB292C54 CRC64;

  Query Match             27.1%;  Score 203.5;  DB 11;  Length 579;
  Best Local Similarity   33.1%;  Pred. No. 3.8e-14;
  Matches   42;  Conservative   22;  Mismatches   56;  Indels    7;  Gaps    4;

Qy       3 SCGSPPPILNGRISYYSTPI-AVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 61
           :|  |||| ||: |      :   |:|: |:|  ||  :|  |::|   ::  | |    |
Db     157 TCEPPPPISNGKHSGRDDDLYTFGSVVIYNCDEPFTLGNASIVCTVVNRTVGVWRPHPP 216

Qy      62 KCEYFNKYSSCPEPIVPGGYKIRGSTP-YRHGDSVTFACKTNFSMNGNKSVWCQANNMWG 120
           |:        |  | :| ||   |    | : |::   ||  | : |:   : |:||  |
Db     217 ACQKI----VCHRPQIPKGYLAPGFRQFYAYRDALEIRCKKGFILRGSSVIHCEANGEWF 272

Qy     121 PTRLPTC 127
           |: :|||
Db     273 PS-IPTC 278


RESULT 21
Q28769
ID   Q28769      PRELIMINARY;      PRT;   322 AA.
AC   Q28769;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR (FRAGMENT).
OS   Papio cynocephalus (Yellow baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RA   Birmingham D.J., Logar C.M., Shen X.-P., Chen W.;
RT   "The baboon erythrocyte complement receptor is a glycophosphatidyl
RT   inositol-linked protein encoded by a homologue of the human CR1-like
RT   genetic element.";
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77977; AAA99004.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 7.
KW   Receptor.
FT   NON_TER       1      1
SQ   SEQUENCE   522 AA;  56626 MW;  312FCB403ADF19DC CRC64;

  Query Match             27.0%;  Score 203;  DB 6;  Length 522;
  Best Local Similarity   38.1%;  Pred. No. 3.8e-14;
  Matches   53;  Conservative   14;  Mismatches   56;  Indels   16;  Gaps    7;

Qy       2 ISCGSPPP LNGRISYYSTP---IAVGTVIRYSC- -SG---TFRLIGEKSLLCITKDKV 52
           | || ||   :|   : ||       |:|: | |   ||    | |:|| |: | :||
Db     169 IPCGLPPT ANG--DFISTSREYFPYGSVVTY CNL GSGRKKLFELVGEPSIYCTSKDDQ 226

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKIR-GST-PYRHGDSVTFACKTNFSMNGNKSV 111
             | |  || :|   ||   | | |  :      :   : | | |:  | | | : |
Db     227 VCIWSCPAPQCIIPNK---CMPPNVENGVLVSYNRS FSLNEVVEFRCQPGFVMKGPRHV 283

Qy     112 WCQANNMWGPTRLPTCVSV 130
             ||| | | |  ||:|  |
Db     284 QCQALNKWEP-ELPSCSRV 301
```

```
RESULT 22
Q08759
ID   Q08759      PRELIMINARY;      PRT:   559 AA.
AC   Q08759;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C4b/C4b RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89010527; PubMed=2971757;
RA   Bourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;
RT   "Identification of an alternative polyadenylation site in the human
RT   C4b/C4b receptor (complement receptor type 1) transcriptional unit and
RT   prediction of a secreted form of complement receptor type 1.";
RL   J. Exp. Med. 168:1255-1270(1988).
DR   EMBL; X14662; CAA32541.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000884; Zn_carbopept.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00436; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN 1.
KW   Signal; Receptor.
FT   NON_TER          1       1
FT   SIGNAL           -1      16        POTENTIAL.
FT   CHAIN            17      559       POTENTIAL.
SQ   SEQUENCE   559 AA;  61424 MW;  DBFFE965CA179D75 CRC64;

  Query Match             26.5%;  Score 199;  DB 4;  Length 559;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-13;
  Matches   52;  Conservative   14;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYSIPTAVGTVIRYSC   SG   TERLIGEKSLLCTIKDKVDG 54
           | ||| || | ||  ||         |:|: | |   ||    | |:|| |: | : |   |
Db     349 IPCGLPPTLINGDFTSTNREYFHYGSVVTYRCNPGSGGRKVFELVGEPSIYCTSNDDQVG 408

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | ||:|   ||   | |  : : :   : | | |: | | | | | | | | | | |
Db     409 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRRVKC 465

Qy     114 QANNMWGPTRLPTCVSV 130
           | | | | |  |:|: |
Db     466 QALNKWEP ELPSCSRV 481


RESULT 23
O08569
ID   O08569      PRELIMINARY;      PRT:   533 AA.
AC   O08569;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ACROSOMAL MATRIX COMPONENT AM67 PRECURSOR.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY; TISSUE=TESTIS;
RX   MEDLINE=97284752; PubMed=9139729;
RA   Foster J.A., Friday B.B., Maulit M.F., Blobel C., Wintrey V.P.,
RA   Olson G.E., Kim K.S., Gerton G.L.;
RT   "AM67, a secretory component of the guinea pig sperm acrosomal matrix,
RT   is related to mouse sperm protein sp56 and the complement component 4
RT   binding proteins.";
RL   J. Biol. Chem. 272:12714-12722(1997).
DR   EMBL; Q75654; AAC13888.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
KW   Signal.
FT   SIGNAL           1      28        POTENTIAL.
FT   CHAIN           29     533        ACROSOMAL MATRIX COMPONENT AM67.
SQ   SEQUENCE   533 AA;  59772 MW;  EDBE2ED487A45389 CRC64;

  Query Match             26.4%;  Score 198.5;  DB 11;  Length 533;
  Best Local Similarity   30.5%;  Pred. No. 1.2e-13;
  Matches   39;  Conservative   26;  Mismatches   56;  Indels    7;  Gaps    4;

Qy       2 ISCGSPPPTLNGRISYYSIPT-AVGTVIRYSCSGTERLIGEKSLLCTIKDKVDGTWDKPA 60
           : | ||| | ||: |     :   |::: | |   : |:|  |: |:  :|  | |
Db     151 VACDSPPDISNGKHSGTDEDLIYYGSLVTYVCDPNYSLIGNASISCTVANKTVGVWSSNP 210

Qy      61 PKCEYFNKYSSCPEPIVPGGYKIRGS-SPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 119
           | ||       | :| :| |  : |    | : |::  :|| : : :: : |:::: |
Db     211 PTCEKV--ICPQPHIPNGSFEYYIYRDILVISCKKGYILEGSSIIRCEANSKW 266

Qy     120 GPTRLPTC 127
             |: :|||
Db     267 YPS IPIC 274


RESULT 24
Q28797
ID   Q28797      PRELIMINARY;      PRT:   522 AA.
AC   Q28797;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MRNA SEQUENCE, 3'END OF ORF (FRAGMENT).
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94292799; PubMed=8021505;
RA   Birmingham D.J., Shen X.P., Bourcade D., Nickells M.W., Atkinson J.P.;
RT   "Primary sequence of an alternatively spliced form of CR1. Candidate
RT   for the 75,000 M(r) complement receptor expressed on chimpanzee
RT   erythrocytes.";
RL   J. Immunol. 153:691-700(1994).
DR   EMBL; L24922; AAA50460.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000884; Zn_carbopept.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00436; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN 1.
FT   NON_TER          1       1
SQ   SEQUENCE   522 AA;  57304 MW;  DB8456EF46200009 CRC64;

  Query Match             26.4%;  Score 198;  DB 6;  Length 522;
  Best Local Similarity   37.2%;  Pred. No. 1.4e-13;
  Matches   51;  Conservative   15;  Mismatches   59;  Indels   12;  Gaps    6;

Qy       2 ISCGSPPPTLNGRISYYSIP-IAVGTVIRYSC   SG  -TERLIGEKSLLCTIKDKVDG 54
           | ||| |    ||  :  |        :|: | |   ||    | |:|| |: | : |   |
Db     106 IICGLPPTLANGDFTSTSREYFHYASVVTYHCNLGSGGKKVFELVGEPSIYCTSKDDQVG 165

Qy      55 TWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGS-TPYRHGDSVTFACKTNFSMNGNKSVWC 113
             | ||:|   ||   | |  : : :   : | | |: | | | | | | | | | | |
Db     166 IWSGPAPQCIIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGPRHVEC 222

Qy     114 QANNMWGPTRLPTCVSV 130
           | | | | |  |:|: |
Db     223 QALNKWEP ELPSCSRV 238
```

```
          || | | |  ||:|  |
Db    224 QALNKWEP ELPSCSRV 238


RESULT 25
Q29531
ID   Q29531      PRELIMINARY;      PRT:   661 AA.
AC   Q29531;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).
GN   CR1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94292799; PubMed=8021505;
RA   Birmingham D.J., Shen X.P., Bourcade D., Nickells M.W., Atkinson J.P.;
RT   "Primary sequence of an alternatively spliced form of CR1. Candidate
RT   for the 75,000 M(r) complement receptor expressed on chimpanzee
RT   erythrocytes.";
RL   J. Immunol. 153:691-700(1994).
DR   EMBL; L24921; AAA51439.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 9.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00042; CCP; 9.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Signal; Receptor.
FT   NON_TER       1      1
FT   SIGNAL        <1     16       POTENTIAL.
FT   CHAIN         17     >661     COMPLEMENT RECEPTOR 1.
FT   NON_TER     661    661
SQ   SEQUENCE   661 AA;  72966 MW;  4F0DF1532664D3F0 CRC64;

  Query Match             26.1%;  Score 196;  DB 6;  Length 661;
  Best Local Similarity   28.0%;  Pred. No. 3.1e-13;
  Matches   56;  Conservative   21;  Mismatches   47;  Indels   76;  Gaps    9;

Qy       2 ISCGSPPPILNGRISYYSTP-- IAVGTVIRYSCSG------TFRLIGEKSLLCITKDKV 52
           | | :|| ||||| :  ||   |  |  | |:|         || |||| |: | :  :
Db     394 IFCPNPPAILNGR--HTGTPFGDIPYGKEISYACDTHPDRGMTFNLIGESSIRCTSDPQG 451

Qy      53 DGTWDKPAPKCEYFNKYSSCPEP---------------IVPG---------GYKIRG--- 85
           :| |  |||:||  :  :: | |                :||             || : |
Db     452 NGVWSSPAPRCE-LSVPAACPHPPKIQNGHDIGGHVSLYLPGMTISYICDPGYLLVGKGF 510

Qy      86 ---------------------STP--------------YRHGDSVTFACKTNFSMNGNK 109
                               | |               | :|| ||  |:  ::: |:
Db     511 IFCTDQGIWSQLDHYCKEVNCSFPLFMNGISKELEMKKVYHYGDYVTLKCEDGYTLEGSP 570

Qy     110 SVWCQANNMWGPTRLPTCVS 129
           |||:: | |  |  | |
Db     571 WSQCQADDRWDPP LAKCTS 589


RESULT 26
Q28770
ID   Q28770      PRELIMINARY;      PRT:   315 AA.
AC   Q28770;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR (FRAGMENT).
OS   Papio cynocephalus (Yellow baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RA   Birmingham D.J., Logar C.M., Shen X.P., Chen W.;
RT   "The baboon erythrocyte complement receptor is a glycophosphatidyl
RT   inositol-linked protein encoded by a homologue of the human CR1-like
RT   genetic element.";
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77978; AAA99005.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     315    315
SQ   SEQUENCE   315 AA;  35172 MW;  305B314AE752E24F CRC64;

  Query Match             25.8%;  Score 194;  DB 6;  Length 315;
  Best Local Similarity   36.7%;  Pred. No. 2.1e-13;
  Matches   51;  Conservative   16;  Mismatches   56;  Indels   16;  Gaps    7;

Qy       2 ISCGSPPPILNGRISYYSTP---IAVGTV RYSC---SG --TFRLIGEKSLLCITKDKV 52
           |||| || | ||   ::|        |:|  |     |     | |:|| |: | : :
Db     123 ISCGLPPTIDNG--DFFSANKEYFHYGSVVTYRCNLGSGGRKLFELVGEPSIYCTSNEDQ 180

Qy      53 DGTWDPPAPKCEYFNKYSSCPEPIVPGGYKIR GSTPYRHGDSVTFACKTNFSMNGNKSV 111
           | |  |||:|   ||   |  | | |  :   : :   : | |  :  | | | : |
Db     181 VGIWSGPAPQCIIPNK---CTPPNVENGI VSVNRSLFSLNEVVEFRCQPGFVMKGPRRV 237

Qy     112 WCQANNMWGPTRLPTCVSV 130
           |||     |  ||:|  |
Db     238 QCQALNKWEP-ELPSCSRV 255


RESULT 27
Q9WRU2
ID   Q9WRU2      PRELIMINARY;      PRT:   645 AA.
AC   Q9WRU2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT BINDING PROTEIN.
OS   Macaca mulatta rhadinovirus 17577.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=85534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99174001; PubMed=10074154;
RA   Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RT   "Sequence and genomic analysis of a rhesus macaque rhadinovirus with
RT   similarity to Kaposi's sarcoma-associated herpesvirus/Human
RT   herpesvirus 8.";
RL   J. Virol. 73:3040-3053(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083501; AAD21332.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
SQ   SEQUENCE   645 AA;  71526 MW;  93D8DBE5ABFB1EB2 CRC64;

  Query Match             25.8%;  Score 194;  DB 12;  Length 645;
  Best Local Similarity   33.1%;  Pred. No. 4.9e-13;
  Matches   44;  Conservative   24;  Mismatches   39;  Indels   25;  Gaps    7;

Qy       4 CGSPPPILNGR-------ISYYSTPIAVGTV RYSCSG FRLIGEKSLLCITKDKVDGTW 56
```

```
                 |  ||: | ||:           |  |        : :||:  | |:|::   ||:       ||
Db    447 CKPEPCIINGKYHPVKDCYQYLDI        VIPSCNKDFSLVGDEMTTCIS      NTW 496

Qy     57 DKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKINFSMNGNKSVWCQA 116
          :||  :||       :|  |  :  |  : :||:  |:|| |||  |:| |:  |::   |:
Db    497 NKPEPRCEQI     ICSAPNIAHGKLLIGSSSVYKYGQSVTIGCETGFTLIGSFISTCK  551

Qy    117 NNMWGPTRLPTCV 128
          :: | |  ||||
Db    552 DSSWDPP LPTCV 563


RESULT 28
Q99JA1
ID   Q99JA1       PRELIMINARY;      PRT;   555 AA.
AC   Q99JA1;
DT   01-JUN-2001 (TrEMBLrel. 12, Created)
DT   01-JUN-2001 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 14, Last annotation update)
DE   C4BP ALPHA CHAIN PRECURSOR.
GN   C4BPA.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPITHELIOMAS, AND LIVER;
RX   MEDLINE=21154058; PubMed=11254714;
RA   Nonaka M.I., Wang G., Mori T., Okada H., Nonaka M.;
RT   "Novel androgen dependent promoters direct expression of the C4b
RT   binding protein alpha chain gene in epididymis.";
RL   J. Immunol. 166:4570-4577(2001).
DR   EMBL; AB049465; BAB39737.1; -.
DR   EMBL; AB049466; BAB39738.1; -.
DR   EMBL; AB049467; BAB39739.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Signal.
FT   SIGNAL        1     13       POTENTIAL.
FT   CHAIN        14    555       POTENTIAL.
SQ   SEQUENCE   555 AA;  61628 MW;  7AFA5462AFC7B6AC CRC64;

  Query Match             25.5%;  Score 191.5;  DB 11;  Length 555;
  Best Local Similarity   33.6%;  Pred. No. 7.8e-13;
  Matches   43;  Conservative   20;  Mismatches   58;  Indels    7;  Gaps    4;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSIICLTKDKVDWTWDKPA 60
           | |  ||| | ||: |     :  |: : |||  :: |:|  |: |   ::  | |
Db     347 IKCKSPEDIRNGKHSGAVFSLKYEGSSVEYSCDEGYSLLGNPSISCTVVNKTVGVWSPSP 406

Qy      61 PRCEYCNKYSSCPEPIVPGGYKIRGSTP YRHGDSVTFACKINFSMNGNKSVWCQANNMW 119
           |:|    |  | :|||     | |  | |   |:: |:|:  | : |:  : |:|:| |
Db     407 EVC    KEVICPQPIVQYAN.ISGFGPTYIEGPIIMESCQGGFVLRGSSLIRCEADNNW 462

Qy     120 GPTRLPTCV 128
           ::  | |
Db     463 NPSP EVC 471


RESULT 29
O02839
ID   O02839       PRELIMINARY;      PRT;   363 AA.
AC   O02839;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PORCINE MEMBRANE COFACTOR PROTEIN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97434414; PubMed=9199970;
RA   Toyomura K., Fujimura T., Murakami H., Natsume T., Shigehisa T.,
RA   Inoue N., Takeda J., Kinoshita T.;
RT   "Molecular cloning of a pig homologue of membrane cofactor protein
RT   (CD46).";
RL   Int. Immunol. 9:869-876(1997).
DR   EMBL; D70897; BAA20476.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   363 AA;  39692 MW;  8A5EF329B8ED4153 CRC64;

  Query Match             25.4%;  Score 190;  DB 6;  Length 363;
  Best Local Similarity   33.6%;  Pred. No. 6.9e-13;
  Matches   45;  Conservative   20;  Mismatches   53;  Indels   16;  Gaps    7;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSC SGT  - ERLIGEKSIICLTKDKVDGT 55
           | |  |  | ||:  : :    |: ||| | |    | |:|| || | | :
Db     169 IICKEPPQETPNGKYSNSHDVFEYNEVVEYSCLSSTGPDEFSLVGESSIFCTKDE     224

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKINFSMNGNKSVWCQ 114
           +    :|:    |   || |:|| |  : :  : :   | | |   |:::|  :: |
Db     225 WSSDPPEC----KVVKCPYPVVPNGEIVSGFGSKFYYKAEVVFKCNAGFTLHGRDLIVCG 280

Qy     115 ANNMWGPTRLPTCV 128
           ||:  |  : | |:
Db     281 ANSTWEP-EMPQCI 293


RESULT 30
Q9DEG0
ID   Q9DEG0       PRELIMINARY;      PRT;   451 AA.
AC   Q9DEG0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT REGULATORY MEMBRANE PROTEIN.
GN   CREMP.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20971924; PubMed=11123420;
RA   Inoue N., Fukui A., Nomura M., Matsumoto M., Toyoshima K., Seya T.;
RT   "A novel chicken membrane associated complement regulatory protein:
RT   Molecular cloning and functional characterization.";
RL   J. Immunol. 166:424-431(2001).
DR   EMBL; AB035592; BAB16878.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
SQ   SEQUENCE   451 AA;  48723 MW;  4FDE12149A502837 CRC64;

  Query Match             25.1%;  Score 188.5;  DB 1;  Length 451;
  Best Local Similarity   32.4%;  Pred. No. 1.3e-12;
  Matches   43;  Conservative   18;  Mismatches   59;  Indels   13;  Gaps    5;

Qy       2 ISCGSPPPILNGRISYYSTPIAVGTVIRYSC------SGTERLICEKSIICLTKDEVDGT 55
           : |  || | |||  :  :   :   | : |||           | :|  |: |   : :|
Db     148 VPCKPPPSIANGRYTEAANYV-YQTTVTYSCDDVRTGENPFSLVGSPSIFCTVDENSNGV 206

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKINFSMNGNKSVWCQ 114
           |  : :|    |   | | | |  | |  : : | :| || | |   :: : |   : |
Db     207 WSGPEPQC----KVVICDNPQVENGKKASGFASQYIYGSSVRFECDPDYVLIGMDVISCI 262
```

```
Qy      115 ANNMWGPTRLPTC 127
            | | |: ||||
Db      263 ENGTWYPS-LPTC 274


RESULT 31
O35520
ID   O35520       PRELIMINARY;       PRT;   417 AA.
AC   O35520;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   512 ANTIGEN (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RT   "Novel isoform of rat complement regulatory protein, rat Crry.";
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D42116; BAA22549.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
FT   NON_TER     417     417
SQ   SEQUENCE   417 AA;  45951 MW;  042D34B68D9E89BDB CRC64;

  Query Match             24.5%;  Score 184;  DB 11;  Length 417;
  Best Local Similarity   33.1%;  Pred. No. 4.7e-12;
  Matches   46;  Conservative   19;  Mismatches   58;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPIILNGRISYYS---TPIAVGTVIRYSCS------GTFRLIGEKSLLCITKDKV 52
           | |  || | ||   ::|        | |: | |:        | |:|| |: | : |
Db      21 IPCEIPPSIPNG--DFFSPNREDFHYGMVVTYQCNTDARGKKLFNLVGEPSIHCTSIDGQ 78

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKI-RGSTPYRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   ||   |  | |     : :  : :   | | | |:  | | |: ||
Db      79 VGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVEFRCQDGFMMKGDSSV 135

Qy     112 WCQANNMWGPTRLPTCVSV 130
           :|:: | | | :||:|  |
Db     136 YCRSLNRWEP-QLPSCFKV 153


RESULT 32
Q63612
ID   Q63612       PRELIMINARY;       PRT;   497 AA.
AC   Q63612;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   512 ANTIGEN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=94161746; PubMed=8117286;
RA   Sakurada C., Seno H., Dohi N., Takizawa H., Nonaka M., Okada N.,
RA   Okada H.;
RT   "Molecular cloning of the rat complement regulatory protein, 512
RT   antigen.";
RL   Biochem. Biophys. Res. Commun. 198:819-826(1994).
DR   EMBL; D42114; BAA07698.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
KW   Signal.
FT   SIGNAL        1     36       POTENTIAL.
FT   CHAIN        37    497       512 ANTIGEN.
SQ   SEQUENCE   497 AA;  54786 MW;  177AC11E3DF1AD1C CRC64;

  Query Match             24.5%;  Score 184;  DB 11;  Length 497;
  Best Local Similarity   33.1%;  Pred. No. 4.6e-12;
  Matches   46;  Conservative   19;  Mismatches   58;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPIILNGRISYYS---TPIAVGTVIRYSCS------GTFRLIGEKSLLCITKDKV 52
           | |  || | ||   ::|        | |: | |:        | |:|| |: | : |
Db     159 IPCEIPPSIPNG--DFFSPNREDFHYGMVVTYQCNTDARGKKLFNLVGEPSIHCTSIDGQ 216

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKI-RGSTPYRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   ||   |  | |     : :  : :   | | | |:  | | |: ||
Db     217 VGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVEFRCQDGFMMKGDSSV 273

Qy     112 WCQANNMWGPTRLPTCVSV 130
           :|:: | | | :||:|  |
Db     274 YCRSLNRWEP-QLPSCFKV 291


RESULT 33
Q63135
ID   Q63135       PRELIMINARY;       PRT;   559 AA.
AC   Q63135;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COMPLEMENT REGULATORY PROTEIN.
GN   CRRY.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=96006570; PubMed=7590969;
RA   Quigg R.J., Lo C.F., Alexander J.J., Sneed A.F., Moxley G. III;
RT   "Molecular characterization of rat Crry: widespread distribution of
RT   two alternative forms of Crry mRNA.";
RL   Immunogenetics 42:362-367(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L36532; AAA91821.1; -.
DR   EMBL; D42115; BAA22548.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
SQ   SEQUENCE   559 AA;  61680 MW;  29E10F6A21DB9B6E CRC64;

  Query Match             24.5%;  Score 184;  DB 11;  Length 559;
  Best Local Similarity   33.1%;  Pred. No. 5.3e-12;
  Matches   46;  Conservative   19;  Mismatches   58;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPIILNGRISYYS---TPIAVGTVIRYSCS------GTFRLIGEKSLLCITKDKV 52
           | |  || | ||   ::|        | |: | |:        | |:|| |: | : |
Db     159 IPCEIPPSIPNG--DFFSPNREDFHYGMVVTYQCNTDARGKKLFNLVGEPSIHCTSIDGQ 216

Qy      53 DGTWDKPAPKCEYFNKYSSCPEPIVPGGYKI-RGSTPYRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   ||   |  | |     : :  : :   | | | |:  | | |: ||
Db     217 VGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVEFRCQDGFMMKGDSSV 273

Qy     112 WCQANNMWGPTRLPTCVSV 130
           :|:: | | | :||:|  |
Db     274 YCRSLNRWEP-QLPSCFKV 291
```

```
RESULT 34
Q64735
ID   Q64735       PRELIMINARY;       PRT:   484 AA.
AC   Q64735;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COMPLEMENT RECEPTOR RELATED PROTEIN.
GN   CRRY
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID-10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN BALB/C;
RX   MEDLINE 90170600; PubMed 2107848;
RA   Earl M.S., Aeqerter Shaw M., Cepok K., Miller M.D., Weis J.H.;
RT   "The murine complement receptor gene family. III. The genomic and
RT   transcriptional complexity of the Cry and Cry ps genes.";
RL   J. Immunol. 144:1088-1096(1990).
DR   EMBL; M44173; AAA37468.1; -.
DR   EMBL; M44164; AAA37468.1; JOINED.
DR   EMBL; M44165; AAA37468.1; JOINED.
DR   EMBL; M44166; AAA37468.1; JOINED.
DR   EMBL; M44167; AAA37468.1; JOINED.
DR   EMBL; M44168; AAA37468.1; JOINED.
DR   EMBL; M44169; AAA37468.1; JOINED.
DR   EMBL; M44170; AAA37468.1; JOINED.
DR   EMBL; M44171; AAA37468.1; JOINED.
DR   EMBL; M44172; AAA37468.1; JOINED.
DR   EMBL; M44173; AAA37467.1; -.
DR   EMBL; M44164; AAA37467.1; JOINED.
DR   EMBL; M44170; AAA37467.1; JOINED.
DR   EMBL; M44171; AAA37467.1; JOINED.
DR   EMBL; M44172; AAA37467.1; JOINED.
DR   EMBL; M44165; AAA37467.1; JOINED.
DR   EMBL; M44166; AAA37467.1; JOINED.
DR   EMBL; M44167; AAA37467.1; JOINED.
DR   EMBL; M44168; AAA37467.1; JOINED.
DR   EMBL; M44169; AAA37467.1; JOINED.
DR   HSSP; P08603; 1HFI.
DR   MGD; MGI:88573; Cry.
DR   InterPro; IPR001005; Myb_DNA-bind.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
DR   PROSITE; PS00037; MYB_1; UNKNOWN 1.
FT   VARIANT       41     83       MISSING (IN LIVER ISOFORM).
SQ   SEQUENCE   484 AA;  53762 MW;  259BFEED4C8547C1 CRC64;

  Query Match             24.0%;  Score 180.5;  DB 11;  Length 484;
  Best Local Similarity   34.6%;  Pred. No. 1.1e-11;
  Matches   47;  Conservative   14;  Mismatches   60;  Indels   15;  Gaps    6;

Qy       2 ISCGSPPPILNGRISYYSI   PIAVGTVIRYSCS --  GPERLIGEKSLLCTIKDKV 52
           | |  || | ||    :::        | ||: | |:          | |:|| || | : |
Db     206 IPCEIPPGIPNG  DFESSTREDFHYGMVVTYRCNTDARGKALFNLVGEPSLYCTSNDGE 263

Qy      53 DgIWDKEAPKCEYENKYSSCPEPIVPGGYKI RGSIPYRHGDSVTFACKTNFSMNGNKSV 111
             | |  | |:|   || :  | | |     :    : :   | | | |   | | |  ||
Db     264 IGVWSGPPPQCIELNKCT  PPPYVENAVMLSENRSLFSLRDIVEFRCHPGFIMKGASSV 321

Qy     112 WCQANNMWGPTRLPTC 127
             ||: | | |  ||:|
Db     322 HCQSLNKWEP ELPSC 336


RESULT 35
Q14006
ID   Q14006       PRELIMINARY;       PRT:   657 AA.
AC   Q14006;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COMPLEMENT B FACTOR (FRAGMENT).
GN   BF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID-9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE 88025472; PubMed 2889480;
RA   Day A.J., Ripoche J., Lyons A., McIntosh B., Harris T.J., Sim R.B.;
RT   "Sequence analysis of a cDNA clone encoding the C terminal end of
RT   human complement factor B.";
RL   Biosci. Rep. 7:201-207(1987).
DR   EMBL; M17517; AAA52016.1; -.
DR   HSSP; P08603; 1HFB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000679; ZnF_GATA.
DR   Pfam; PF00084; sushi; 11.
DR   PRINTS; PR00619; GATAZNFINGER.
DR   SMART; SM00032; CCP; 11.
FT   NON_TER       1       1
SQ   SEQUENCE   657 AA;  74247 MW;  6862CDC5239F1213 CRC64;

  Query Match             24.0%;  Score 180.5;  DB 4;  Length 657;
  Best Local Similarity   34.8%;  Pred. No. 1.6e-11;
  Matches   47;  Conservative   23;  Mismatches   50;  Indels   15;  Gaps    8;

Qy       3 SCGSPPPILNGRISYYSIPIAVGT-IVIRYSCSGIPERLIGEKSLLCTIKDKVIP-IWDKPA 60
           ||| || :||| :    |   |    | : | | :  | : |    :  |     || : |
Db      55 SCGPPPELLNGNVK-EKTKEEYGHSEVVEYYCNPRFLMKGPNKIQC    VDGEW TTL 107

Qy      61 PKCPYEPEYSSCPE-PIVPGGYEIRGSIPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 119
           | |    : |:| : | :  |:    | || :|||| | |  :|:  ,:::: ,  : :|
Db     108 PVC -IVEESTGGDIPELEHGWAQLSSPPYYYGDSVEFNCSESFIMIGHRSTIC  IHGVW 164

Qy     120 GDTRLPTCVSVFPLE 134
             |:|| ||::  |:
Db     165 --TQLPQCVAIDKLK 177


RESULT 36
Q91X48
ID   Q91X48       PRELIMINARY;       PRT:   469 AA.
AC   Q91X48;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT COMPONENT 4 BINDING PROTEIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID-10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE-LIVER;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC012257; AAH12257.1; -.
SQ   SEQUENCE   469 AA;  51523 MW;  14266B414430F7107 CRC64;

  Query Match             24.8%;  Score 178.5;  DB 11;  Length 469;
  Best Local Similarity   34.3%;  Pred. No. 1.8e-11;
  Matches   45;  Conservative   13;  Mismatches   54;  Indels   23;  Gaps    7;

Qy       2 ISCGSPPPIINGRIS--  -YYSTPIAVGTVIRYSCSGTPRLIGEKSLLCTIKDKVIDGTWD 57
           : || || | :|:| :     :|  |   |  | |:|   |||:|   : |   :|   |
Db     179 VKCGPPPDISNGKHSGTEDFY--PYNHG  ISYICDPGFRLVGSPSIGCIVVNKIVDVWS 234

Qy      58 KPAPKCEYENEYSSCPEP     IVPGGYEIRGSIPYRHGDSVTFACKTNFSMNGNKSVW 112
             | |       | :|     ::  |||      | | |||  ||     : |  :
```

```
Db    285 SSPPTCEKI----ICSQPNILHGVIVSGYK----ATYTHRDSVRLACLNGTVLRGRHVIE 286

Qy    113 CQANNMWGPTRLPTC 127
          || |  |  : ||||
Db    287 CQGNGNW--SSLPTC 299


RESULT 37
P97263
ID   P97263      PRELIMINARY;      PRT;   337 AA.
AC   P97263;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN PRECURSOR.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY; TISSUE=TESTIS;
RA   Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT   "Molecular cloning of guinea pig membrane cofactor
RT   protein:preferential expression in testis.";
RL   J. Immunol. 0:0-0(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY; TISSUE=TESTIS;
RX   MEDLINE=97098723; PubMed=8943400;
RA   Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT   "Molecular Cloning of Guinea Pig Membrane Cofactor
RT   Protein:preferential expression in testis.";
RL   J. Immunol. 157:4946-4952(1996).
DR   EMBL; D84132; BAA12234.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Signal.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    337       MEMBRANE COFACTOR PROTEIN (GM3).
SQ   SEQUENCE   337 AA;  37377 MW;  0AFC4DFFC4D1FDBB CRC64;

  Query Match             23.4%;  Score 176;  DB 11;  Length 337;
  Best Local Similarity   30.2%;  Pred. No. 2.2e-11;
  Matches   42;  Conservative   21;  Mismatches   60;  Indels   16;  Gaps    6;

Qy      2 ISCGSPPPILNGRISYYSTPI-AVGTVIRYSCSGT     FRLIGEKSLLCITKDKVDGT 55
          : |  || | ||: ::    :      : |||         |:| : | |    |
Db    160 VLCSPPPKIQNGKYTFSDVQVFEYFEAVTYSCDAVQGPDKLSLVCNEVLYCACHQK---- 215

Qy     56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
          |   ||:|    |   || :|  | :| |    : :  :||| |   |  ||: :| |
Db    216 WSSAAPEC----KVVKCPLPVVKNGKQISGLGQTFFYQATVTFQCLPGFYFNGSSTVVCG 271

Qy    115 ANNMWGPTRLPTCVSVFPL 133
          ::| | |: :| |: |  |
Db    272 SDNIWKPS IPECLKVVGL 289


RESULT 38
Q9N0B7
ID   Q9N0B7      PRELIMINARY;      PRT;  1172 AA.
AC   Q9N0B7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DJ177P10.1.1 (H FACTOR 1 (COMPLEMENT) ISOFORM 1).
GN   HF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bird C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL049744; CAB70597.1; -.
DR   HSSP; P08603; 1HFH.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 19.
DR   SMART; SM00032; CCP; 19.
SQ   SEQUENCE   1172 AA;  132087 MW;  EF54054C484FA454 CRC64;

  Query Match             23.4%;  Score 176.5;  DB 4;  Length 1172;
  Best Local Similarity   35.7%;  Pred. No. 1.1e-10;
  Matches   46;  Conservative   20;  Mismatches   48;  Indels   15;  Gaps    8;

Qy      3 SCGSPPPILNGRISYYSTPIAVG--TVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPA 60
          ||| || :||| :    |    |   |: | |:  | : |   : |     ||| |
Db    629 SCGPPPELINGNVK-EKTKEEYGHSEVVEYYCNPRFLMKGPNKIQC---VDGEW-TIL 681

Qy     61 PKCEYFNKYSSCPE-PIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANNMW 119
          | |    : |:| : | :  |:    | :| :|||| | |  :|:| |:: : |  : :|
Db    682 PVC--IVEESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCSESFTMIGHRSTIC IHGVW 738

Qy    120 GPTRLPTCV 128
            :|| ||
Db    739 --TQLPQCV 745


RESULT 39
Q9R0R9
ID   Q9R0R9      PRELIMINARY;      PRT;   300 AA.
AC   Q9R0R9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SOLUBLE FORM OF MEMBRANE COFACTOR PROTEIN CD46.
GN   MCP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20093984; PubMed=10630285;
RA   Nomura M., Tsujimura A., Shida K., Matsumoto M., Matsuda Y.,
RA   Toyoshima K., Seya T.;
RT   "Membrane and secretory forms of murine membrane cofactor protein
RT   (CD46) generated from a single gene through alternative splicing.";
RL   Immunogenetics 50:245-254(1999).
DR   EMBL; AB022600; BAA86056.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:1203290; Mcp.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   300 AA;  33658 MW;  1B57EA1655EA1F14 CRC64;

  Query Match             22.9%;  Score 172;  DB 11;  Length 300;
  Best Local Similarity   32.6%;  Pred. No. 5.4e-11;
  Matches   45;  Conservative   17;  Mismatches   56;  Indels   20;  Gaps    7;

Qy      2 ISCGSPPPILNGRISYYSTPIAV   CTVIEYSCSGT-----FRLIGEKSLLCITKDKVD 53
          | |   | | ||  ::  | | |       : |||  |     | |:|   : |   :
Db    171 IYCLPPPKIKNG--THTLTDINVEKYHEAVSYSCDTTPGPDKFSLVGTSMIECAGHN  225

Qy     54 GTWDKPAPKCEYFNKYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVW 112
           ||    |:|    |   || |:  |  | |:   : :  :| | |   | : |:  |
Db    226 -TWSNSPPEC----KVVKCPNPVLQNGRIIGAGETFSYQSTVMFECLQGFYMEGSSMVI 280

Qy    113 CQANNMWGPTRLPTCVSV 130
          | ||| | |: :| |: |
Db    281 CSANNSWEPS-IPKCLKV 297
```

RESULT 40
P97939
ID P97939 PRELIMINARY; PRT; 267 AA.
AC P97939;
DT 01 MAY 1997 (TrEMBLrel. 03, Created)
DT 01 MAY 1997 (TrEMBLrel. 03, Last sequence update)
DT 01 DEC 2001 (TrEMBLrel. 19, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN MCP.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=LIVER;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular cloning of guinea pig membrane cofactor
RT protein preferential expression in testis.";
RL J. Immunol. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=LIVER;
RX MEDLINE=97098723; PubMed=8944400;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular cloning of Guinea Pig Membrane Cofactor
RT Protein:preferential expression in testis.";
RL J. Immunol. 157:4946-4952(1996).
DR EMBL; D84134; BAA12236.1; -.
DR EMBL; D86339; BAA12236.1; JOINED.
DR EMBL; D86340; BAA12236.1; JOINED.
DR EMBL; D86341; BAA12236.1; JOINED.
DR EMBL; D86342; BAA12236.1; JOINED.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
FT NON_TER 1 1
FT NON_TER 267 267
SQ SEQUENCE 267 AA; 29753 MW; 659E2EC15CE3E3BC CRC64;

Query Match 22.8%; Score 171; DB 11; Length 267;
Best Local Similarity 29.9%; Pred. No. 6e-11;
Matches 40; Conservative 21; Mismatches 57; Indels 16; Gaps 6;

Qy 2 ISCGSPPPIINGRISYYSTPI-AVGTVIRYSCSGT-- PFILIGEKSLICITKDKVDGT 55
: | || | | :: : : ||| | :|: : | | |
Db 126 VLCSPPPKIQNGKYTFSDVQVFEYFEAVTYSCDAVQGPDKLSIVGNFVIYCAGHQK-- - 181

Qy 56 WDKEAPKCEYFNKYSSCPEPIVPGGYKIRG SIPYRHGDSVIFACKTNFSMNGNKSVWCQ 114
| || :| | | || :| | :| | : : :|| | | || :| |
Db 182 WSSAAPEC KVVKCPLPVVKNGKQISGLQIFEYQAIVIFQCLPGFYFNGSSTVVCG 237

Qy 115 ANNMWGPTRLPTCV 128
:| | | :| |
Db 238 SDNIWKPS IPECL 250

RESULT 41
P97261
ID P97261 PRELIMINARY; PRT; 339 AA.
AC P97261;
DT 31 MAY 1997 (TrEMBLrel. 03, Created)
DT 01 MAY 1997 (TrEMBLrel. 03, Last sequence update)
DT 01 DEC 2001 (TrEMBLrel. 19, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN PRECURSOR (FRAGMENT).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular cloning of guinea pig membrane cofactor protein:
RT preferential expression in testis.";
RL J. Immunol. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RX MEDLINE=97098723; PubMed=8944400;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular Cloning of Guinea Pig Membrane Cofactor
RT Protein:preferential expression in testis.";
RL J. Immunol. 157:4946-4952(1996).
DR EMBL; D88358; BAA14594.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular cloning of guinea pig membrane cofactor
RT protein:preferential expression in testis.";
RL J. Immunol. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RX MEDLINE=97098723; PubMed=8944400;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular Cloning of Guinea Pig Membrane Cofactor
RT Protein:preferential expression in testis.";
RL J. Immunol. 157:4946-4952(1996).
DR EMBL; D84130; BAA12232.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Signal.
FT NON_TER 1 1
FT SIGNAL -1 19 POTENTIAL.
FT CHAIN 20 339 MEMBRANE COFACTOR PROTEIN (GMI).
SQ SEQUENCE 339 AA; 37870 MW; 1461C309F76F02CF CRC64;

Query Match 22.8%; Score 171; DB 11; Length 339;
Best Local Similarity 29.9%; Pred. No. 8.1e-11;
Matches 40; Conservative 21; Mismatches 57; Indels 16; Gaps 6;

Qy 2 ISCGSPPPIINGRISYYSTPI-AVGTVIRYSCSGT- PFILIGEKSLICITKDKVDGT 55
: | || | | :: : : ||| | :|: : | | |
Db 145 VLCSPPPKIQNGKYTFSDVQVFEYFEAVIYSCDAVQGPDKLSIVGNFVIYCAGHQK 200

Qy 56 WDKEAPKCEYFNKYSSCPEPIVPGGYKIRG SIPYRHGDSVIFACKTNFSMNGNKSVWCQ 114
| || :| | | || :| | :| | : : :|| | | || :| |
Db 201 WSSAAPEC---KVVKCPLPVVKNGKQISGLQIFEYQAIVIFQCLPGFYFNGSSTVVCG 256

Qy 115 ANNMWGPTRLPTCV 128
:| | | :| |
Db 257 SDNIWKPS IPECL 269

RESULT 42
P70105
ID P70105 PRELIMINARY; PRT; 379 AA.
AC P70105;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN(GMPI-FULL) PRECURSOR.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular cloning of guinea pig membrane cofactor protein:
RT preferential expression in testis.";
RL J. Immunol. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=TESTIS;
RX MEDLINE=97098723; PubMed=8944400;
RA Hosokawa M., Nonaka M., Okada N., Nonaka M., Okada H.;
RT "Molecular Cloning of Guinea Pig Membrane Cofactor
RT Protein:preferential expression in testis.";
RL J. Immunol. 157:4946-4952(1996).
DR EMBL; D86339; BAA12236.1; -.
DR HSSP; P10998; 1VVD.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.

```
DR   SMART; SM00032; CCP; 4.
KW   Signal.
FT   SIGNAL         1     34       POTENTIAL.
FT   CHAIN         35    379       POTENTIAL.
SQ   SEQUENCE   379 AA;  42159 MW;  5DA56B058C8A1573 CRC64;

  Query Match             22.8%;  Score 171;  DB 11;  Length 379;
  Best Local Similarity   29.9%;  Pred. No. 9.2e-11;
  Matches   40;  Conservative   21;  Mismatches   57;  Indels   16;  Gaps    6;

Qy       2 ISCGSPPPEILNGRISYYSTPI AVGTVIRYSCSGT-----FRLIGEKSLLCITKDKVDGT 55
           : |  || | ||: ::    :      : |||          |:| : | |    |
Db     160 VICSPPPKIQNGKYTFSDVQVFEYFEAVTYSCDAVQGPDKLSLVGNEVLYCAGHQK---- 215

Qy      56 WDKPAPKCEYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKTNFSMNGNKSVWCQ 114
           |   ||:|    |   || |:|  | :| |    : :  .||| |   |  ||. .| |
Db     216 WSSAAPEC----KVVKCPIPVVKNGKQISGLGQTFFYQATVTFQCLPGFYFNGSSTVVCG 271

Qy     115 ANNMWGPTRLPTCV 128
           ::| | |: :| |:
Db     272 SDNTWKPS-IPECL 284


RESULT 43
Q9QAX6
ID   Q9QAX6      PRELIMINARY;      PRT;   268 AA.
AC   Q9QAX6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   YB-C5R.
GN   YB-C5R.
OS   Yaba monkey tumor virus
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Yatapoxvirus.
OX   NCBI_TaxID=38804;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Amano H., Kato K., Miyamura T.;
RT   "Sequence analysis of Yaba virus DNA.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB025319; BAA88977.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   268 AA;  30606 MW;  4EF681CE5A93D6EC CRC64;

  Query Match             22.6%;  Score 169.5;  DB 12;  Length 268;
  Best Local Similarity   41.3%;  Pred. No. 8.9e-11;
  Matches   41;  Conservative   21;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       8 SCGSPPEILNGRISYYSTPIAVGTVIRYSCS----GTFRLIGEKSLLCITKDKVDGTWDK 58
           :|  |  |||| |      ::|  : | |          : |:| ::: ||     || |
Db      79 NCKDPVMILNGNIKNKQDKYSLGDSVIYMCKVNKLERYSLVGNETVKCI-----DGKWVP 133

Qy      59 PAPKCEYFN-KYSSCPEPIVPGGYKIRGSTPYRHGDSVTFACKTNFSMNGNKSVWCQANN 117
             | |:    || :    |: | :| :     : :|| | | ||| | ::|: :   |  |
Db     134 DNPFCKLIRCKYPALQNGILEGSFKKK----FYYGDVVIFKCKTGFRLSGSLTSTCGINF 189

Qy     118 MWGPTRLPTCV 128
           :| |  || ||
Db     190 VWVP-NLPKCV 199


RESULT 44
PB8903
ID   PB8903      PRELIMINARY;      PRT;   550 AA.
AC   PB8903;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF 4.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97094384; PubMed=8939871;
RA   Moore P.S., Boshoff C., Weiss R.A., Chang Y.;
RT   "Molecular mimicry of human cytokine and cytokine response pathway
RT   genes by KSHV.";
RL   Science 274:1739-1744(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97121480; PubMed=8962146;
RA   Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RT   "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RT   (HHV8).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Russo J.J., Bohenzky R.A., Chien M.-C., Chen J., Yan M., Maddalena D.,
RA   Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75698; AAC57082.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   550 AA;  60688 MW;  D4B8B2B4BAC1CCB5 CRC64;

  Query Match             22.6%;  Score 169.5;  DB 12;  Length 550;
  Best Local Similarity   33.3%;  Pred. No. 2.1e-10;
  Matches   42;  Conservative   14;  Mismatches   51;  Indels   19;  Gaps    7;

Qy       8 PPILNGRI----SYYSTPIAVGTVIRYSCSGIFRLIGEKSLLCITKDKVDGIWDKPAPKC 63
           | | ||       ||         : | :  : |:|  |: |      |: ||    | |
Db     156 PKIKNGDFKPDKDYYE----YNDAVHFECNEGYTLVGEHSIAC--  --AVNNTWISNMPIC 207

Qy      64 EYFNKYSSCPEPIVPGGYKIRG-STPYRHGDSVTFACKINFSMNGNKSVWCQANNMWGPT 122
           |      |  | |  || |:| |  |:  ||||||   | : |: ::      | |
Db     208 E----LAGCKFPSVTHGYPIQGFSLIYKHKQSVTFACNDGFVLRGSPTITCNVIE WDPP 262

Qy     123 RLPTCV 128
             || ||
Db     263 -LPKCV 267


RESULT 45
O40912
ID   O40912      PRELIMINARY;      PRT;   550 AA.
AC   O40912;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF 04.
OS   Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=37296;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97296220; PubMed=9151804;
RA   Neipel F., Albrecht J.C., Fleckenstein B.;
RT   "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
RT   human herpesvirus 8: determinants of its pathogenicity?";
RL   J. Virol. 71:4187-4192(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
RA   Friedman-Kien A.E., Fleckenstein B.;
RT   "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
```

DB EMBL; D93872; AAB02602.1; -.
DB BSSP; P10998; IVVD.
DB InterPro; IPR000436; Sushi SCR CCP.
DB Pfam; PF00084; sushi; 4.
DB SMART; SM00032; CCP; 4.
SQ SEQUENCE 550 AA; 60648 MW; 8A7EC1C30E79C6A6 CRC64;

Query Match 22.4%; Score 168.5; DB 12; Length 550;
Best Local Similarity 33.3%; Pred. No. 2.7e-10;
Matches 42; Conservative 14; Mismatches 51; Indels 19; Gaps 7;

```
Qy        8 PPILNGRI     SYYSTPIAVGTVIRYSCNSGTERLIGEKSLICITKDKVDGTWDKPAPKC 63
            | | ||        ||          : : |:   : |:|  |: |      |: ||    | |
Db      156 PKIENGDFKPDKDYYE     YNDAVHFECNEGYTLVGPHSIAC----AVNNTWISNMPTC 207

Qy       64 EYPNKYSSCPEPIVPGGYKIRG STPYRHGDSVTFACKTNFSMNGNKSVWCQANNMWGPT 122
            |      : |  | |  || : :| |  |:|  |||||||  | : |: : : |     | |
Db      208 P     LAGCKFPSVTHGYPTQGFSLIYKHKQSVTFACNDGFVLRGSPTITCNVIE-WDPP 262

Qy      123 RLPICV 128
             || ||
Db      263 LPKCV 267
```

Search completed: November 6, 2002, 16:08:22
Job time : 27.4201 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 27.1004 Seconds
(without alignments)
553.312 Million cell updates/sec

Title: US-09-834-309-6
Perfect score: 752
Sequence: 1 EISCDPPPEVKNARKPYYSL.....ANEMWGPTALPVCESDFPLE 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_042802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	466.5	62.0	1087	12	AAR11139	B lymphocyte membr
2	225.5	30.0	581	12	AAR13490	Human C4 binding p
3	221.5	29.5	2039	14	AAR36743	CR1. Homo sapiens
4	220.5	29.3	1542	12	AAR11982	Partial human comp
5	220.5	29.3	1929	22	ABG00103	Novel human diagno
6	220.5	29.3	1930	19	AAW45899	Human complement r
7	220.5	29.3	2039	12	AAR11810	Human complement f
8	220.5	29.3	2039	20	AAY55751	Human C3b/C4b rece
9	220.5	29.3	2039	22	ABC00287	Novel human diagno
10	220.5	29.3	2044	22	ABB11782	Human CR1 protein
11	220.5	29.3	2044	22	AAM39224	Human polypeptide
12	220.5	29.3	2044	22	AAM41010	Human polypeptide
13	220.5	29.3	2317	10	AAP92219	CR1 protein. Homo
14	215	28.6	579	19	AAW39924	Amino acid sequenc
15	212.5	28.3	450	20	AAY55755	Human CR1 protein
16	204.5	27.2	646	20	AAY55758	Human CR1 protein
17	199.5	26.5	363	18	AAW12414	Porcine complement
18	199.5	26.5	363	20	AAY30918	MCP protein. Unid
19	197.5	26.3	263	20	AAY29858	Vaccinia virus SPI
20	197.5	26.3	263	20	AAY29860	Mutated VCP giving
21	195.5	26.0	263	22	AAB48846	Cowpox virus infla
22	193.5	25.7	263	10	AAP92003	Deduced sequence o
23	193.5	25.7	263	20	AAY29859	Vaccinia complemen
24	193.5	25.7	263	21	AAB13014	Complement inhibit
25	189.5	25.2	324	12	AAR15233	CD46 from clone pm
26	189.5	25.2	336	12	AAR15232	CD46 from clone pm
27	189.5	25.2	350	12	AAR15231	CD46 from clone pm
28	189.5	25.2	357	12	AAR10927	Human membrane cof
29	189.5	25.2	373	12	AAR15230	CD46 from clone pm
30	189.5	25.2	377	12	AAR15229	CD46 from clone pm
31	188.5	25.1	251	14	AAR41461	Membrane cofactor
32	188.5	25.1	254	17	AAW06880	Membrane co factor
33	188.5	25.1	279	14	AAR41460	Membrane cofactor
34	188.5	25.1	279	19	AAW69523	rSMCP protein SEQ
35	188.5	25.1	376	17	AAR93939	CD46 wild type H
36	188.5	25.1	377	17	AAR93942	CD46 construct sub
37	188.5	25.1	377	17	AAR93941	CD46 construct sub
38	188.5	25.1	377	18	AAW27484	Human MCP. Homo s
39	188.5	25.1	378	17	AAR93940	CD46 construct sub
40	188.5	25.1	384	12	AAR10924	Human membrane cof
41	188.5	25.1	384	16	AAR86416	Human CD46. Homo
42	188.5	25.1	421	21	AAR58794	Lung cancer associ
43	188.5	25.1	421	22	AAG75528	Human colon cancer
44	188.5	25.1	577	17	AAW06882	Membrane cofactor
45	188.5	25.1	611	22	AAE12569	CAB2 protein. Uni
46	188.5	25.1	611	22	AAE03762	CAB 2 chimeric pro
47	187.5	24.9	314	17	AAR93943	CD46 construct del
48	185.5	24.7	645	21	AAB53125	Macaca mulatta rha
49	178.5	23.7	133	15	AAR47156	Sequence of solubl
50	178.5	23.7	254	15	AAR47154	Sequence of solubl
51	178.5	23.7	438	20	AAY55756	Human CR1 protein
52	178.5	23.7	450	20	AAY55753	Human CR1 protein
53	178.5	23.7	496	20	AAY55752	Human CR1 protein
54	178.5	23.7	778	19	AAW73147	Amino acid sequenc
55	175.5	23.3	254	15	AAR47155	Sequence of solubl
56	175.5	23.3	481	13	AAR29092	CR1-4 (amino acids
57	175.5	23.3	483	13	AAR29091	CR1-4 (amino acids
58	175.5	23.3	543	13	AAR28543	CR1-4 (35E, 37Y) a
59	175.5	23.3	543	13	AAR28544	CR1-4 (35E) analog
60	175.5	23.3	543	13	AAR28545	CR1-4 (37Y) analog
61	175.5	23.3	543	13	AAR28546	CR1-4 (44T, 47L) a
62	175.5	23.3	543	13	AAR28547	CR1-4 (52S, 54S, 5
63	175.5	23.3	543	13	AAR28548	CR1-4 (57V, 59E) a
64	175.5	23.3	543	13	AAR28549	CR1-4 (64K, 65L) a
65	175.5	23.3	543	13	AAR28550	CR1-4 (64K) analog
66	175.5	23.3	543	13	AAR28551	CR1-4 (65T) analog
67	175.5	23.3	543	13	AAR28552	CR1-4 (78T, 79G) a
68	175.5	23.3	543	13	AAR28553	CR1-4 (85R, 87N) a
69	175.5	23.3	543	13	AAR28554	CR1-4 (92L, 94H) a
70	175.5	23.3	543	13	AAR28555	CR1-4 (92L) analog
71	175.5	23.3	543	13	AAR28556	CR1-4 (94H) analog
72	175.5	23.3	543	13	AAR28557	CR1-4 (99H, 103E)
73	175.5	23.3	543	13	AAR28558	CR1-4 (109N, 111A,
74	175.5	23.3	543	13	AAR28559	CR1-4 (114 117S,KP
75	175.5	23.3	543	13	AAR28560	CR1-4 (114S) analo
76	175.5	23.3	543	13	AAR28561	CR1-4 (115L) analo
77	175.5	23.3	543	13	AAR28562	CR1-4 (116K) analo
78	175.5	23.3	543	13	AAR28563	CR1-4 (117E) analo
79	175.5	23.3	543	13	AAR28564	CR1-4 (116K, 117P)
80	175.5	23.3	543	13	AAR28565	CR1-4 (121Q) analo
81	175.5	23.3	543	13	AAR28566	CR1-4 (418R, 419N)
82	175.5	23.3	543	13	AAR28567	CR1-4 (418 421 RNP
83	175.5	23.3	543	13	AAR28568	CR1-4 (447T, 449Y)
84	175.5	23.3	543	13	AAR28569	CR1-4 (469 476 STK

85	175.5	23.8	544	1	AAR28570	CR1 4 (266-274 KLK
86	175.5	23.8	544	1	AAR28571	CR1 4 (364-467 NAA
87	169	22.5	367	15	AAR50087	MCP:DAF fusion pro
88	168.5	22.4	450	20	AAY55754	Human CR1 protein
89	168.5	22.4	453	20	AAY55757	Human CR1 protein
90	164.5	21.9	1653	22	ABG71860	Drosophila melanog
91	163.5	21.7	764	20	AAW94376	Human Factor B ana
92	162.5	21.6	300	22	AAB61858	Myxoma virus immun
93	162.5	21.6	764	20	AAW94380	Human Factor B ana
94	162	21.5	920	22	ABG05921	Novel human diagno
95	162	21.5	920	22	ABG22404	Novel human diagno
96	161	21.4	792	21	AAB42725	Human ORFX ORF2489
97	161	21.4	982	22	ABG01221	Novel human diagno
98	161	21.4	990	22	ABG05922	Novel human diagno
99	161	21.4	990	22	ABG22405	Novel human diagno
100	161	21.4	996	22	ABG05925	Novel human diagno

ALIGNMENTS

```
RESULT 1
AAR11139
ID   AAR11139 standard; Protein  1087 AA.
XX
AC   AAR11139;
XX
DT   23-MAY-1991  (first entry)
XX
DE   B-lymphocyte membrane glycoprotein CR2.
XX
KW   CR2; B-lymphocyte; membrane receptor protein; Epstein Barr virus;
KW   extracellular domain.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          20..90
FT                   /note= "1"
FT   Region          20..154
FT                   /note= "2"
FT   Region          20..276
FT                   /note= "3"
FT   Region          20..652
FT                   /note= "4"
FT   Region          20..776
FT                   /note= "5"
FT   Region          20..1025
FT                   /note= "6"
FT   Peptide         1..20
FT                   /label= sig_peptide
FT   Protein         21..1087
FT                   /label= mat_protein
XX
PN   WO9102546-A.
XX
PD   07-MAR-1991.
XX
PF   22-AUG-1990;   90WO-US004817.
XX
PR   23-AUG-1989;   89US-0398224.
XX
PA   (SCRI ) SCRIPPS CLINIC & RES.
XX
PI   Moore MD;  Cooper ND;  Nemerow GR;
XX
DR   WPI; 1991-087109/12.
DR   N-PSDB; AAQ10989.
XX
PT   Synthetic polypeptide(s) of extracellular domain - treat
PT   Epstein-Barr virus infections and diagnose same by formation of
PT   complex between CR2 and EBV
XX
PS   Disclosure; Fig 9.1-9.9; 60pp; English.
XX
CC   The six indicated fragments encode claimed and disclosed
CC   polypeptides, which are synthesised by recombinant expression, pref.
CC   in a baculovirus expression system in which a DNA plasmid, contg. a
CC   cDNA encoding this sequence, is truncated to encode a CR2 polypeptide
CC   that comprises a region of the extracellular domain of CR2.  The
CC   construct is then inserted downstream of an appropriate promoter in
CC   a transfer vector and integrated into a baculovirus which is then
CC   used to infect host insect cells for expression.
CC   The polypeptides correspond to B-lymphocyte membrane receptor
CC   protein for Epstein-Barr Virus (EBV) and give specific binding.
CC   Protein number 3 is used to inhibit infection of mammalian cells in
CC   contact with an aq. medium, (esp. mammalian blood) and of human
CC   B-lymphocytes, by EBV.  The peptides are used to detect the presence
CC   of EBV in an aq. sample and to detect antibodies directed against CR2.
CC   The peptides are used in a pharmacological compsn. as active
CC   ingredient with a carrier to treat immune disorders.
XX
SQ   Sequence   1087 AA;

  Query Match             62.0%;  Score 466.5;  DB 12;  Length 1087;
  Best Local Similarity   61.2%;  Pred. No. 6.7e-38;
  Matches   82;  Conservative   18;  Mismatches   44;  Indels    1;  Gaps    1;

Qy       2 ISCPPPPEVKNARKEYYYSIPIVFGIVIFYIGSSVYLSGPEFAIFCISENQVRRAIWQEAPP 61
           |||  |  : | |  ||| ||  |||:|||:|| :|||||||:: ||:::::  |||  |
Db      21 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSLLCITKDKVDGTWDKPAP 80

Qy      62 IKESVNKTIFSPEIPVSCPMNRSSFAHLRRDSVIRLRANEIMEGKEVWGLANSMWG 121
           ||  |   || :|||||:  :|:  |:|||||||| ||  |:::  :::|||| | |||
Db      81 KCEYFNKYSSCPEPIVPGGYKIRGS IPYRHGDSVTFACKINFSMNGNKSVWCQANNMWG 140

Qy     122 PTALPVCESDFPLE 135
           || ||     .|:|
Db     140 PTRLPTCVSVFPLE 153


RESULT 2
AAR13490
ID   AAR13490 standard; Protein; 581 AA.
XX
AC   AAR13490;
XX
DT   30-OCT-1991  (first entry)
XX
DE   Human C4 binding protein.
XX
KW   C4bp, monomer; complement protein; pDOL:C4bp.4; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..32
FT                   /label= signal_peptide
FT   Protein         33..581
FT                   /label= C4bp
FT   Region          33..93
FT                   /label= SCR8
FT   Region          94..155
FT                   /label= SCR7
FT   Region          156..219
FT                   /label= SCR6
FT   Region          220..279
FT                   /label= SCR5
FT   Region          280..345
FT                   /label= SCR4
FT   Region          346..406
FT                   /label= SCR3
FT   Region          407..464
FT                   /label= SCR2
FT   Region          465..523
```

```
FT                   /label= SCR1
FT   Domain          524..581
FT                   /label= C4bp_core
FT                   /note= "responsible for multimer assembly"
FT   Disulfide-bond  34..80
FT                   /note= "intradomain"
FT   Disulfide-bond  65..92
FT                   /note= "intradomain"
XX
PN   WO9111461 A.
XX
PD   08 AUG 1991.
XX
PF   28 JAN 1991;    91WO-US00567.
XX
PR   26 JAN 1990;    90US-0470888.
XX
PA   (BIOG ) BIOGEN INC.
XX
PI   Pasek MP,  Winkler G,  Liu TK;
XX
DR   WPI; 1991-252613/34.
DR   N-PSDB; AAQ13242.
XX
PT   New C4 binding protein fusion proteins and DNA encoding them -
PT   comprise assemblies of C4bp monomers linked to functional moiety,
PT   e.g. AZ1, useful as delivery vehicles in diagnosis and therapy
XX
PS   Example 1; Fig 1; 105pp; English.
XX
CC   This sequence was deduced from human hepatocyte (Hep G2) cDNA
CC   obtained following PCR amplification. The protein is a monomer
CC   containing 8 SCRs. Each SCR forms a looped domain due to the
CC   presence of two intradomain disulphide bonds (only the disulphide
CC   bonds of SCR8 are labelled in the Features Table). Within each SCR,
CC   the first cysteine residue bonds with the third and the second
CC   cysteine residue bonds with the fourth. This secondary structure is
CC   responsible for the conformational flexibility of the C4bp monomer.
CC   The invention covers fusion proteins in which the monomer sequence,
CC   or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)
CC   is fused to the C-terminal of a protein such as a viral receptor,
CC   cell ligand, a bacterial, viral or parasitic immunogen, enzyme,
CC   cytokine, toxin, etc.  See also AAQ13243-51.
XX
SQ   Sequence   581 AA;

  Query Match             30.0%;  Score 225.5;  DB 12;  Length 581,
  Best Local Similarity   31.8%;  Pred. No. 3.9e-14;
  Matches   42;  Conservative   25;  Mismatches   52;  Indels   13;  Gaps    4;

Qy       2 ISCDPPPEVKNAR---KPYYSLPIVPGTIVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           : | |||:::| |   : : :     |  : |:| | : |:|  :| |  ||:    |
Db     157 VKCKPPPDIRNGRHSGEENFYAY    GFSVTYSCDPRFSLLGHASISCTVENETIGVWR 212

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             :|| ||     |:|  | |  | |   : : |:: | |:  | ::|| : | |:
Db     213 PSPPTCEK----IICKKPDVSHGEMVSGFGPIYNYKDTIVFKCQKGFVLRGSSVIHCDAD 268

Qy     118 EMWGPTALPVCE 129
             | |:  | ||
Db     269 SKWNPSPPACE 279


RESULT 3
AAR36743
ID   AAR36743 standard; Protein; 2039 AA.
XX
AC   AAR36743;
XX
DT   14 SEP 1994  (first entry)
XX
DE   CR1.
XX
KW   C3b/C4b receptor; CR1; erythrocyte; monocyte; macrophage; granulocyte;
KW   iC3b; T cell; splenic follicular dendritic cell; soluble; complement;
KW   glomerular podocyte; B cell; C3b; C4b; inactivated C3b; phagocytosis;
KW   plasma; ligand binding activity; immune complex; activator; allotype;
KW   endocytosis, lymphocyte, classical, alternative, pathway, cofactor; F;
KW   C3/C5 convertase; liver; cleavage; factor I; regulation; glycoprotein;
KW   S; A; B; glycosylation; duplication; repetitive intervening sequence;
KW   endoglycosidase F.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..40
FT                   /note= "Signal peptide"
FT   Protein         503..2039
FT                   /note= "CR1 encoded by CR1 cDNA"
FT   Region          503..490
FT                   /note= "Long homologous repeat-A"
FT   Region          491..1393
FT                   /note= "Long homologous repeat-B"
FT   Region          1394..1843
FT                   /note= "Long homologous repeat-C"
FT   Misc-difference 161..162
FT                   /note= "Unclear in specification"
XX
PN   US5212071-A.
XX
PD   18-MAY-1993.
XX
PF   01-APR-1988;    88US-0176532.
XX
PR   01-APR-1988;    88US-0176532.
PR   03-APR-1989;    89US-0332865.
XX
PA   (BCHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (TCEL-) T CELL SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Carson GR,  Concino MF,  Fearon DT,  Ip SH,  Klickstein LB;
PI   Makrides SC,  Wong WW;
XX
DR   WPI; 1993-175454/21.
DR   N-PSDB; AAQ41867.
XX
PT   Nucleic acid encoding polypeptide having complement regulatory
PT   activity - used to prevent reperfusion injury, inhibit Arthus
PT   reaction and neutrophil mediated tissue damage and reduce
PT   myocardial infarct size and inflammation
XX
PS   Claim 1; Fig 1; 90pp; English.
XX
CC   This sequence represents the entire amino acid sequence of the C3b/C4b
CC   receptor (CR1).  CR1 is present on erythrocytes, monocytes/macro-
CC   phages, granulocytes, B cells, some T cells, splenic follicular
CC   dendritic cells and glomerular podocytes.  CR1 specifically binds
CC   C3b, C4b and inactivated C3b (iC3b).  A soluble form of the receptor
CC   is found in plasma which has ligand binding activity and the same
CC   molecular weight as membrane-associated CR1.  CR1 binds C3b and C4b
CC   that have covalently attached to immune complexes and other complement
CC   activators.  The consequences of these interactions depends on the
CC   type of bearing the receptor.  Erythrocyte CR1 binds immune complexes
CC   for transport to the liver.  CR1 on neutrophils and monocytes
CC   internalises bound complexes, either by adsorptive endocytosis
CC   or by phagocytotsis.  The function of CR1 on B lymphocytes is less
CC   well defined.  CR1 can inhibit the classical and alternative pathway
CC   C3/C5 convertases and act as a cofactor for the cleavage of C3b and
CC   C4b by factor I, therefore CR1 has a complement regulatory function
CC   as well as acting as a receptor.  CR1 is a glycoprotein composed of
CC   a single polypeptide chain.  Four allotypic forms of CR1 have been
CC   identified, differing by increments of approx. 40-50 kD.  The two most
CC   common forms, the F and S allotypes, also termed A and B allotypes,
CC   have molecular weights of 250 and 290 kD respectively.  The two rarer
CC   forms have molecular weights of 210 and 290 kD.  These differences
```

```
CC   represent differences in the polypeptide chain of CR1, rather than
CC   glycosylation state because they are not abolished by treatment of
CC   purified receptor portein with endoglycosidase F.  The CR1 gene has
CC   been shown to have repetitive intervening sequences which may have
CC   been duplicated in the formation of the larger allotypes.
XX
SQ   Sequence   2039 AA;

  Query Match             29.5%;  Score 221.5;  DB 14;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 4.5e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:|||  : |   : :||        |||: | |          :::|:|||:|:| |::
Db    1517 ISCEPPPTISN--RDFYSNNRTSFHNGTVVTYQCHTGPDGEQLFQLVGERSIYCTSKDDQ 1574

Qy      53 HATWDKAPPTCESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFIMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |   | | || ||
Db    1575 VGVWSSPPPRCISTNK--CTAPEVQNAIRVPGNRSFFSLTFIIFRFCHPGPVMVGSHTV 163

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| |: |
Db    1632 QCQTNGRWGP KLPHC 1646


RESULT 4
AAR11982
ID   AAR11982 standard; Protein; 1537 AA.
XX
AC   AAR11982;
XX
DT   25-JUN-1991  (first entry)
XX
DE   Partial human complement type 1 receptor.
XX
KW   complement system; C3b/C4b receptor; CR1; allergic reaction;
KW   immune response; long homologous repeat; LHR.
XX
OS   Homo sapiens
XX
FH   Key             Location/Qualifiers
FT   Region          1..438
FT                   /label= LHR B
FT   Region          439..891
FT                   /label= LHR C
FT   Region          892..1443
FT                   /label= LHR D
FT   Region          1495..1498
FT                   /note= "positively charged; preceded by hydrophobic
FT                   sequence"
FT   Region          1521..1526
FT                   /note= "has 67 per cent homology to site of protein
FT                   kinase C phosphorylation in the EGF
FT                   receptor"
XX
PN   WO9105047-A
XX
PD   18-APR-1991.
XX
PF   25-SEP-1990;  90WO-US05454.
XX
PR   26-SEP-1989;  89US-0412745.
PR   26-SEP-1990;  90US-0587459.
XX
PA   (CYCE-) T CELL SCI INC.
PA   (JOHN-) JOHNS HOPKINS UNIVERSITY.
PA   (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI   Fearon DT;  Klickstein LB;  Wong WW;  Carson GR;  Hoh M;  Concino MF;
PI   Makrides SC;  Marsh HC;
XX
DR   WPI; 1991-120854/18.
DR   N-PSDB; AAQ11643.

XX
FT   Human complement receptor type 1 gene, encoded proteins and
PT   fragments - for treatment of immune disorders, myocardial infarct,
PI   damage due to inflammmation and in treatment of thrombosis
XX
PS   Disclosure; Fig 5; 234pp; English.
XX
CC   This sequence comprises three of the four tandem, direct, long
CC   homologous repeats of the full length F allozyme of CR1. LHR A is
CC   absent. Each LHR might represent a single C3b/C4b binding domain,
CC   making the receptor multivalent. The LHR's are composed of 7 short
CC   consensus repeats of 60-70 residues resembling the SCR's of other
CC   C3/C4 binding proteins. The protein and fragments of it having C4b
CC   and/or C4b binding activity can be used to treat immune disorders
CC   or disorders involving inappropriate complement activity.
CC   See also AAQ11642.
XX
SQ   Sequence   1537 AA;

  Query Match             29.3%;  Score 220.5;  DB 12;  Length 1537;
  Best Local Similarity   36.0%;  Pred. No. 4e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YPLIGEKAIFCISENQV 52
           |||:|||  : |   :||        |||: | |          : :|:|||:|  ::
Db    1015 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGFQLFFLVGERSIYCTSKDDQ 1072

Qy      53 HATWDKAPPTCESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFIMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |   | | || ||
Db    1073 VGVWSSPPPRCISTNK-  CTAPEVENAIRVPGNRSFFSLTFIIFRFCQPGPVMVGSHTV 1129

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| |: |
Db    1130 QCQTNGRWGP KLPHC 1144


RESULT 5
ABG00103
ID   ABG00103 standard; Protein; 1929 AA.
XX
AC   ABG00103;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #94.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639462/73.
DR   N-PSDB; AAS64290.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 30462; 104pp; English.
```

```
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010 ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1929 AA;

  Query Match             29.4%;  Score 220.5;  DB 22;  Length 1929;
  Best Local Similarity   36.0%;  Pred. No. 5.3e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLERYTCSPS------YRLIGEKAIFCISENQV 52
           |||:|||  .|    :||        |||: | |         : |:|||:|:| |::
Db    1407 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1464

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFPHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1465 VGVWSSPPPRCISINK--GTAPEVENAIRVPGNRSFFSITEIIRFRCQPGFVMVGSHTV 1521

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1522 QCQTNGRWGP-KLPHC 1576


RESULT 6
AAW45899
ID   AAW45899 standard; peptide; 1930 AA.
XX
AC   AAW45899;
XX
DT   30-JUN 1998  (first entry)
XX
DE   Human complement receptor 1 (residues 1-1929).
XX
KW   Membrane binding element; thrombotic disease; soluble protein;
KW   complement-related disease; integral membrane protein; inflammation;
KW   short consensus repeat; SCR 1-3; CR1; complement receptor type 1.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Cross-links     1930
FT                   /note= "Disulphide linked to Cys in peptide given
FT                          in AAW45889"
XX
PN   WO9802454 A2.
XX
PD   22-JAN-1998.
XX
PF   08-JUL-1997;   97WO-EP03715.
XX
PR   15-JUL-1996;   96GB-0014871.
XX
PA   (ADPR-) ADPROTECH PLC.
XX
PI   Dodd I, Mossakowska DEI, Smith RAG;
XX
DR   WPI; 1998-110524/10.
XX
PT   Derivatives of soluble poly:peptide(s) bonded to low affinity
PT   membrane binding groups - useful for treating complement-related and
PT   thrombotic diseases, providing improved localisation at cellular
PT   membranes
XX
PS   Claim 22; Pages 60-61; 75pp; English.
XX
CC   This sequence represents human complement receptor 1 (CR1, CD 35)
CC   N-terminal fragment. The invention relates to a soluble derivative (A)
CC   of a soluble polypeptide (I), which comprises at least 2 heterologous
CC   membrane-binding elements (MBE) of low membrane affinity covalently
CC   associated with (I). MBE interact, independently and with thermodynamic
CC   additivity, with components of cellular or artificial membranes exposed
CC   to extracellular fluids. (A) are used to treat disorders treatable with
CC   (I) itself, specifically inflammation or any other complement related
CC   disorder (e.g. neurological disease, graft rejection, myocardial
CC   infarction, sepsis, rheumatoid arthritis and many others; including
CC   application to indwelling devices) and thrombolytic disease, but also to
CC   treat allergy, induce weight loss, to treat ischaemia or asthma and as
CC   immuno-modulators for treating multiple sclerosis. (A) are administered
CC   orally, topically, by injection or inhalation at 0.01-10 (preferably
CC   0.1-10) mg/kg/day.
XX
SQ   Sequence   1930 AA;

  Query Match             29.3%;  Score 220.5;  DB 19;  Length 1930;
  Best Local Similarity   36.0%;  Pred. No. 5.3e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLERYTCSPS------YRLIGEKAIFCISENQV 52
           |||:|||  .|    :||        |||: | |         : |:|||:|:| |::
Db    1476 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1533

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFPHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1534 VGVWSSPPPRCISINK---GTAPEVENAIRVPGNRSFFSITEIIRFRCQPGFVMVGSHTV 1590

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1591 QCQTNGRWGP-KLPHC 1605


RESULT 7
AAR11810
ID   AAR11810 standard; Protein; 2039 AA
XX
AC   AAR11810;
XX
DT   25-JUN-1991  (first entry)
XX
DE   Human complement type 1 receptor.
XX
KW   complement system; C3b/C4b receptor; CR1; allergic reaction;
KW   immune response; clone lambda T109.1
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..41
FT                   /label= putative signal peptide
FT   Protein         42..2039
FT                   /label= CR1
XX
PN   WO9105047 A.
XX
PD   18-APR-1991.
XX
PF   25-SEP-1990;   90WO-US05454.
XX
```

```
FF   26-SEP-1989;   89US-0412345.
FF   26-SEP-1990;   90US-0912349.
XX
PA   (SCPI ) I CELL SCI INC.
PA   (UYJO ) JOHNS HOPKINS UNIVERSITY.
PA   (BRIG ) BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI   Fearon DT,  Klickstein LB,  Wong WW,  Carson GR,  Hoh M,  Concino MF;
PI   Makrides SC,  Marsh HC;
XX
DR   WPI; 1991-132854/18.
DR   N-PSDB; AAQ11642.
XX
PT   Human complement receptor type 1 gene, encoded proteins and
PT   fragments - for treatment of immune disorders, myocardial infarct,
PT   damage due to inflammation and in treatment of thrombosis
XX
PS   Claim 41; Fig 1; 244pp; English.
XX
CC   The invention also covers fragments of this protein which have the
CC   ability to bind C3b and/or C4b, have cofactor I activity or can
CC   inhibit C3 or C5 convertase activity. The full-length protein, or
CC   its specified fragments are used to treat patients with immune
CC   disorders or a disorder caused by inappropriate complement
CC   activity. The protein is also used to treat thrombotic conditions
CC   in humans and animals. See also AAQ11644.
XX
SQ   Sequence  2039 AA;

  Query Match             29.4%;  Score 220.5;  DB 12;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 5.6e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDEPPEVENARRFYYS--LDIVPCTVLFYLCSPS      YRLICFKKAIEFISENQV 52
           |||.||| : |    :||        ||| | |         : |:|||:|:| |::
Db    1517 ISCEEPPLISNG-DFYSNNRISFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 HAIWPKAPFICESVNKIISCSPFIVPGGFMNRSGKAFEFIEFCKANFIMFGSFTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  :| | || ||
Db    1575 VGVWSSPPPRCISINK-CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy     113 WCQANEMWGPIALPVC 128
           || |  ||| || |
Db    1632 QCQTNGRWGP-KLPHC 1646


RESULT 8
AAY55751
ID   AAY55751 standard; Protein; 2039 AA.
XX
AC   AAY55751;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human C3b/C4b receptor (CR1) protein.
XX
KW   C3B/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; diagnostic.
XX
OS   Homo sapiens.
XX
PN   EP1001481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470552.
XX
PR   28-APR-1989;   89US-0343865.
PR   06-DEC-1974;   74US-0530238.
PR   24-FEB-1993;   93US-0026434.
PR   01-APR-1988;   88US-0176532.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BCHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-633852/54.
DR   N-PSDB; AAZ48150.
XX
PT   A human C4B/C4B receptor (CR1) protein having anti-inflammatory and
PT   cardiant activity -
XX
PS   Disclosure; Fig 1A-P; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the human CR1 protein.
XX
SQ   Sequence  2039 AA;

  Query Match             29.4%;  Score 220.5;  DB 20;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 5.6e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDEPPEVENARRFYYS--LDIVPCTVLFYLCSPS      YRLICFKKAIEFISENQV 52
           |||.||| : |    :||        ||| | |         : |:|||:|:| |::
Db    1517 ISCEEPPLISNG-DFYSNNRISFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 HAIWPKAPFICESVNKIISCSPFIVPGGFMNRSGKAFEFIEFCKANFIMFGSFTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  :| | || ||
Db    1575 VGVWSSPPPRCISINK-CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy     113 WCQANEMWGPIALPVC 128
           || |  ||| || |
Db    1632 QCQTNGRWGP-KLPHC 1646


RESULT 9
ABG00287
ID   ABG00287 standard; Protein; 2039 AA.
XX
AC   ABG00287;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #278.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
```

XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/74.
DR N-PSDB; AAS64474.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 30646; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2039 AA;

Query Match 29.3%; Score 220.5; DB 22; Length 2039;
Best Local Similarity 36.0%; Pred. No. 5.6e-13;
Matches 49; Conservative 18; Mismatches 54; Indels 15; Gaps 5;

Qy 2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS-----YRLIGEKAIFCISENQV 52
DB 1517 ISCEPPPTISNG--DFYSNNKTSFHNGIVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy 53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
DB 1575 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy 113 WCQANEMWGPTALPVC 128
DB 1632 QCQTNGRWGP-KLPHC 1646

RESULT 10
ABB11782
ID ABB11782 standard; peptide; 2044 AA.
XX
AC ABB11782;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human CR1 protein homologue, SEQ ID NO:2152.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration, wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; carciant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 04-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR N-PSDB; ABA09026.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 20; Page 215-246, 1963pp, English.
XX
CC Sequences ABB10981 ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 2044 AA;

Query Match 29.3%; Score 220.5; DB 22; Length 2044;
Best Local Similarity 36.0%; Pred. No. 5.7e-13;

```
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS----YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : |:|||:|| |::
Db    1522 ISCEPPPFISNG--DFYSNNRTSFHNGTVVYTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |        |::: |   : : | |:  | | || ||
Db    1580 VGVWSSPPPRCISINK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1637 QCQTNGRWGP-KLPHC 1651


RESULT 11
AAM40224
ID   AAM40224 standard; Protein; 2044 AA.
XX
AC   AAM40224;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2369.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens
XX
PN   WO200153312 A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0633450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE ) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AA158380.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2369; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AA157798 AA161369) and
CC   the encoded polypeptides (AAM38642 AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy Drager Syndrome. Other uses include the
CC   utilisation of the activities such as, immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   2044 AA;

  Query Match             29.3%;  Score 220.5;  DB 22;  Length 2044
  Best Local Similarity   36.0%;  Pred. No. 5.7e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS----YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : |:|||:|| |::
Db    1522 ISCEPPPFISNG--DFYSNNRTSFHNGTVVYTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |        |::: |   : : | |:  | | || ||
Db    1580 VGVWSSPPPRCISINK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1637 QCQTNGRWGP-KLPHC 1651


RESULT 12
AAM41010
ID   AAM41010 standard; Protein; 2044 AA.
XX
AC   AAM41010;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 5941.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312 A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0633450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AA160166.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 5941; 10078pp; English.
```

```
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   2044 AA;

  Query Match             29.4%;  Score 220.5;  DB 22;  Length 2044;
  Best Local Similarity   36.0%;  Pred. No. 5.7e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||         ||| | |          : |:|||:::| |::
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1637 QCQTNGRWGP-KLPHC 1651


RESULT 13
AAP92219
ID   AAP92219 standard; protein; 2317 AA.
XX
AC   AAP92219;
XX
DT   22-FEB-1990  (first entry)
XX
DE   CR1 protein.
XX
KW   Complement; cofactor.
XX
OS   Homo sapiens (human).
XX
FH   Key             Location/Qualifiers
FT   Peptide         10..50
FT                   /label= signal_peptide
XX
PN   WO8909220-A.
XX
PD   05-OCT-1989.
XX
PF   31-MAR-1989;   89WO-US01358.
XX
PR   01-APR-1988;   88US-0176532.
XX
PA   (TCEL ) T CELL SCIENCES INC.
PA   (CYJO ) THE JOHNS HOPKINS UNIVERSITY.
PA   (BRIG ) THE BRIGHAM AND WOMEN'S HOSPITAL.
XX
PI   Fearon DT,  Klickstein LB,  Wong W,  Carson G,  Concino MF,  Makrides SC;
XX
DR   WPI; 1989-309498/42.
DR   N-PSDB; AAN91477.
XX
PT   New nucleic acid sequences encoding new CR1 protein - and its fragment,
PT   for diagnosis and control of complement-related immune defects,
PT   inflammation, myocardial infarct, etc
XX
PS   Claim 1; Fig 1; 191pp; English.
XX
CC   This is full-length CR1 protein, and shortened forms are new, lacking
CC   the transmembrane region.  The proteins and fragments bind C4b and/or
CC   C4b, have cofactor activity and inhibit C3 and C5 convertase activity
CC   In the sequence, x-untranslated region.  This has 7 short consensus
CC   repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in
CC   patent).  Deletion mutants, eg LHR-A for binding of C4b, and LHR-C
CC   for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.
CC   They are useful in diagnosing and treating immune disorders, and prevent
CC   perfusion injury.
XX
SQ   Sequence   2317 AA;

  Query Match             29.3%;  Score 220.5;  DB 10;  Length 2317;
  Best Local Similarity   36.0%;  Pred. No. 6.1e-13;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||         ||| | |          : |:|||:::| |::
Db    1526 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1583

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1584 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1640

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1641 QCQTNGRWGP-KLPHC 1655


RESULT 14
AAW39924
ID   AAW39924 standard, Protein; 579 AA.
XX
AC   AAW39924;
XX
DT   19-JUN-1998  (first entry)
XX
DE   Amino acid sequence of a mouse sperm protein designated sp56.
XX
KW   Sperm protein; sp56; protein receptor; Sushi domain;
KW   species restricted epitope; sp56 antigenic determinant;
KW   immunocontraceptive; vaccine; sperm fertility; sp56 antigen.
XX
FH   Key             Location/Qualifiers
FT   Domain          33..91
FT                   /note= "Sushi domain 1"
FT   Domain          96..153
FT                   /note= "Sushi domain 2"
FT   Domain          158..250
FT                   /note= "Sushi domain 3"
FT   Domain          223..278
FT                   /note= "Sushi domain 4"
FT   Domain          283..345
FT                   /note= "Sushi domain 5"
FT   Domain          345..411
FT                   /note= "Sushi domain 6"
FT   Domain          45..509
FT                   /note= "Sushi domain 7"
FT   Domain          412..487
FT                   /note= "sp56 specific sequence domain"
FT   Domain          415..430
FT                   /note= "sp56 specific sequence domain"
FT   Domain          428..443
FT                   /note= "sp56 specific sequence domain"
FT   Domain          438..453
FT                   /note= "sp56 specific sequence domain"
XX
OS   Mus sp.
```

```
XX
PN   WO9604411-A1.
XX
PD   08-FAN-1996
XX
PF   03-AUG-1995;   95WO-US11219.
XX
PR   03-AUG-1994;   94US-0285895.
XX
PA   (SCRI-) SCRIPPS RES INST.
XX
PI   Bleil JD,  Bookbinder LH,  Wolin MB;
XX
DR   WPI; 1996-088608/08.
DR   N-PSDB; AAV09901.
XX
PT   New immunocontraceptive vaccine compositions - containing an
PT   antigenic polypeptide having an sp56 antigenic determinant to induce
PT   anti-sp56 antibodies in mammals
XX
PS   Claim 6; Pages 65-97; 129pp; English.
XX
CC   The present sequence represents a murine sperm protein designated sp56.
CC   The primary sequence of the sp56 protein indicates that it is a
CC   member of a superfamily of protein receptors, characterised by the
CC   presence of a series of 60 amino acid long homology repeats called Sushi
CC   domains. Each Sushi domain contains 10 virtually invariant amino acids.
CC   A stretch of 44 amino acids near the carboxy terminus is believed to
CC   define a species restricted epitope. sp56 antigenic determinants are used
CC   in an immunocontraceptive vaccine composition. The vaccine composition
CC   induces anti sp56 antibodies in a mammalian subject, in combination with
CC   an excipient. The compositions can be used for oral immunocontraception
CC   for mammals. The method can also be used for assessing sperm fertility,
CC   detection of sperm sp56 antigens and confirming effectiveness of the
CC   vaccines at inducing immune responses. The immunocontraceptive can be
CC   formulated to contain species restricted sp56 sequences such that
CC   contraception is only achieved in the target species, and not in other
CC   animals or humans who inadvertently consume the bait.
XX
SQ   Sequence   579 AA;

  Query Match             28.6%;  Score 215;  DB 19;  Length 579;
  Best Local Similarity   40.3%;  Pred. No. 4.4e-14;
  Matches   40;  Conservative   27;  Mismatches   51;  Indels   14;  Gaps    4;

Qy       5 SCDPPEEVKNAR      KPYYSLEPIVPGTVLRYRCSPSYRLIGEKAIFCISENQVHWTWD 57
           :: ||| : | :          |       |:|: | | |:: |:|  :| |   |:    |
Db     157 DCPPPPPISNGKRSGRDHDLEYE    GSVVIYNCDPIFTLLGNASIVCIVVNRTVGVWR 212

Qy      58 KAPFICESVNKTISCSDPTVPGGEMNKGSKAPFRHGDSVIFTCKANFIMKGSKIVWCQAN 117
             || |:     | |  | :| |::  | :  : : |::   || | ::|| : |:||
Db     213 PHPPACQK     IVCHRPQTPKGYIAPGFRQFYAYRDALEIRCKKGFILRGSSVIHCEAN 268

Qy     118 IMWGPIALPVCE 129
             | | ::| ||
Db     269 CPWFP SIPDCE 279


RESULT 15
AAY55759
ID   AAY55759 standard; Protein  450 AA.
XX
AC   AAY55759;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR D SCR fragment.
XX
KW   C4b/C4B receptor; CR1 protein; cell surface protein; erythrocyte; human;
KW   complement regulatory activity; complement pathway enzyme; tissue damage;
KW   reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW   heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
KW   short consensus repeat.
XX
OS   Homo sapiens.
XX
PN   US5981481-A.
XX
PD   09-NOV-1999.
XX
PF   06-JUN-1995;   95US-0470652.
XX
PR   03-APR-1989;   89US-0332865.
PR   06-DEC-1974;   74US-0450238.
PR   24-FEB-1993;   93US-0026134.
PR   01-APR-1988;   88US-0176532.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI   Concino MF,  Wong WW,  Makrides SC,  Klickstein LB,  Fearon DT,  Ip SH;
PI   Marsh HC,  Carson GR;
XX
DR   WPI; 1999-633357/54.
XX
PT   A human C3B/C4B receptor (CR1) protein having anti-inflammatory and
PT   cardiant activity
XX
PS   Disclosure; Fig 10; 87pp; English.
XX
CC   The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC   protein or fragment is expressed as a cell surface protein on the surface
CC   of a non-human cell and exhibits a complement regulatory activity of full
CC   -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC   may be mediated through the inhibition of complement pathway enzymes. The
CC   soluble CR1 protein exhibits a complement regulatory activity, and this
CC   may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC   neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC   and inflammation. The CR1 protein and its fragments can also be used in
CC   the treatment of conditions which involve unwanted complement activity,
CC   e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC   and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC   -CR1 antibodies are used in assays, and diagnostics. The present sequence
CC   represents the short consensus repeat (SCR) fragments of human CR1
CC   protein long homologous repeat (LHR) D sequence.
XX
SQ   Sequence   450 AA;

  Query Match             28.3%;  Score 212.5;  DB 20;  Length 450;
  Best Local Similarity   45.3%;  Pred. No. 5.8e-14;
  Matches   48;  Conservative   18;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       2 ISCDPPEEVKNARKPYYS-- LPIVPGIVLRYTCSPS       YRLIGEKAIFCISENQV 52
           |||:||| : |    :||          |||: | |           : :|:|::|:| |::
Db     124 ISCEPPFIISNG--DFYSNNGTSFHNGTVVIYQCHTGPDGEQLFELVGERSIYCTSKDDQ 181

Qy      53 HATWDKAPFICESVNKTISCSDPTVPGGEMNKGSKAPFRHGDSVIFTCKANFIMKGSKIV 112
             |   || | | ||   |:   |       |::: .   : : | |:    ' '| ||
Db     182 VGVWSSPPFRCISTNK---CTAENVENAIRVPGNRSFTSLIEIIRFRC-PYPVMVGSHIV 238

Qy     113 WCQANEMWGPIALPVC 128
             || |  |||  || |
Db     239 QCQTNGRWGP-KLPHC 253


RESULT 16
AAY55758
ID   AAY55758 standard; Protein; 646 AA.
XX
AC   AAY55758;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Human CR1 protein LHR-D fragment.
XX
```

KW C3B/C4B receptor; CR1 protein; cell-surface protein; erythrocyte; human;
KW complement regulatory activity; complement pathway enzyme; tissue damage;
KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
KW heart condition; autoimmune disorder; long homologous repeat; LHR
XX
OS Homo sapiens.
XX
PN US5981481 A.
XX
PD 09 NOV 1999.
XX
PF 06 JUN 1995; 95US-0470652.
XX
PR 03 APR 1989; 89US-0332865.
PR 06 DEC 1974; 74US-0530238.
PR 24 FEB 1993; 93US-0026134.
PR 01 APR 1988; 88US-0176532.
XX
PA (DYJO) UNIV JOHNS HOPKINS.
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (AVAN) AVANT IMMUNOTHERAPEUTICS INC.
XX
PI Concino MF, Wong WW, Makrides SC, Klickstein LB, Fearon DT, Ip SH;
PI Marsh HC, Carson GR;
XX
DR WPI; 1999-633857/54.
XX
PT A human C3B/C4B receptor (CR1) protein having antiinflammatory and
PT cardiant activity -
XX
PS Disclosure; Fig 5B; 87pp; English.
XX
CC The invention relates to a human C3B/C4B receptor (CR1) protein. The CR1
CC protein or fragment is expressed as a cell-surface protein on the surface
CC of a non-human cell and exhibits a complement regulatory activity of full
CC -length human CR1 as expressed on erythrocytes. The CR1 function in vivo
CC may be mediated through the inhibition of complement pathway enzymes. The
CC soluble CR1 protein exhibits a complement regulatory activity, and this
CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC and inflammation. The CR1 protein and its fragments can also be used in
CC the treatment of conditions which involve unwanted complement activity,
CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC and autoimmune disorders. CR1 proteins, analogues, derivatives, and anti
CC CR1 antibodies are used in assays, and diagnostics. The present sequence
CC represents the human CR1 protein long homologous repeat (LHR)-D fragment.
XX
SQ Sequence 646 AA;

Query Match 27.2%; Score 204.5; DB 20; Length 646;
Best Local Similarity 35.3%; Pred. No. 5.6e-12;
Matches 48; Conservative 16; Mismatches 57; Indels 15; Gaps 5;

Qy 2 ISCDPPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
Db 124 ISCEPPP--TSANCDEYSSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGRPSIYCTSKDDQ 181

Qy 54 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
Db 182 VGVWSSPPPRCISINK--CIAPEVENAIRVPGNRFLFSLTEIIRFRCQLPGFVMVGSHTV 238

Qy 113 WCQANEMWGPTALPVC 128
Db 239 QCQTNGRWGP-KLPHC 253

RESULT 17
AAW12414
ID AAW12414 standard; Protein; 363 AA.
XX
AC AAW12414;
XX
DT 24-SEP-1997 (first entry)
XX
DE Porcine complement inhibitor.
XX
KW porcine, pig; complement; inhibitor, organ transplantation;
KW analysis; promoter.
XX
OS Sus scrofa.
XX
PN WO9700951-A1.
XX
PD 09-JAN-1997.
XX
PF 19-JUN-1996; 96WO-JP01704.
XX
PR 20-JUN-1995; 95JP-0178254.
XX
PA (NIME-) NIPPON MEAT PACKERS INC.
PA (NIHA-) NIPPON HAM KK.
XX
PI Murakami H, Shigehisa T, Toyomura K;
XX
DR WPI; 1997-087378/08.
DR N-PSDB; AAT61098.
XX
PT DNA encoding porcine complement inhibitor - useful in porcine organ
PT transplant to humans
XX
PS Claim 3; Page 12-14; 20pp; Japanese.
XX
CC This protein is a porcine complement inhibitor encoded by pMCPcDNA
CC (AAT61098). The DNA is useful for large scale production of
CC recombinant porcine complement inhibitor, which is useful for
CC porcine organ transplantation into humans. The DNA clone pMCPcDNA is
CC also useful in the analysis of the promoter region of porcine complement
CC inhibitor.
XX
SQ Sequence 363 AA;

Query Match 26.5%; Score 199.5; DB 14; Length 363;
Best Local Similarity 31.9%; Pred. No. 8.4e-12;
Matches 45; Conservative 21; Mismatches 52; Indels 23; Gaps 6;

Qy 1 EISCDPPPEVKNAR-----KPYYSLPIVPGTVLRYTCSPS--- -YRLIGEKAIFCISEN 50
Db 168 KILCKPPGEIPNGKYTNSHKDVFEY----NEVVIYSCLSSTGPDEFSLVGESSLFCIGKD 223

Qy 51 QVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSK 110
Db 224 E----WSSDPPEC----KVVKCPYPVVPNGEIVSGFGSKFYYKAEVVFKCNAGFTLHGRD 275

Qy 111 TVWCQANEMWGPTALPVCESD 131
Db 276 TIVCGANSTWEP-EMPQCIKD 295

RESULT 18
AAY30918
ID AAY30918 standard; Protein; 363 AA.
XX
AC AAY30918;
XX
DT 18-OCT-1999 (first entry)
XX
DE MCP protein.
XX
KW MCP; reporter gene; complement inhibitor; expression; antibody;
KW monoclonal antibody; IgG1; extracellular.
XX
OS Unidentified.
XX
PN JP11206392-A.
XX
PD 03-AUG-1999.

XX
PT 28-SEP-1998; 98GB-0201555.
XX
PR 26-SEP-1997; 97JP-0259547.
XX
PA (NIHA) NIPPON HAM KK.
XX
DR WPI; 1999-488062/41.
DR N-PSDB; AAZ09145.
XX
PT New reporter gene useful for distinguishing extracellularly the
PT presence and/or extent of expression of gene
XX
PS Disclosure; Page 13-14; 18pp; Japanese.
XX
CC This invention describes a novel reporter gene, MCP, consisting of a
CC gene of complement inhibitor. The invention also describes a method for
CC the measurement of expression of MCP in which a gene containing the above
CC MCP gene (reporter gene) is introduced to a cell derived from an animal
CC species other than the animal species derived from the above MCP gene to
CC prepare a transformant and then an antibody recognizing MCP is reacted to
CC measure MCP. The invention also describes a monoclonal antibody of IgG1
CC type recognizing pig complement inhibitor MCP (pMCP). The method can
CC distinguish extracellularly the presence and/or the extent of expression
CC of a gene. This sequence represents the MCP protein of the invention.
XX
SQ Sequence 363 AA;

Query Match 26.5%; Score 199.5; DB 20; Length 363;
Best Local Similarity 41.9%; Pred. No. 8.9e-12;
Matches 45; Conservative 21; Mismatches 52; Indels 23; Gaps 8;

Qy 3 ETSCDPPPEVKNAR KPYYSLPIVIGIVLRYTCSPS 50
Db 168 KTTCKPPQGFTPNGKYINSHKDVFEY NEVVIYSCLSSFGPDEPSLVCESSLFCTGKD 225

Qy 51 QVHATWDKAPPICESVNKTTSCSDFIVPGGEMNKGSKAPFRHGDSVIFICKANFIMKGSK 110
Db 226 E WSSDPTEC KVVKCTYTVVINGEIVSGEGSKEYYKAEVVFKCNAGFTLHGE 275

Qy 111 IVWCQANEMWGPTALPVCESD 131
Db 276 IIVCGANSTWEP EMPQCIKD 295

RESULT 19
AAY29858
ID AAY29858 standard; Protein; 263 AA.
XX
AC AAY29858;
XX
DT 16-NOV-1999 (first entry)
XX
DE Vaccinia virus SPICE protein
XX
KW Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW fusion protein; hyperacute rejection; xenograft; inflammation;
KW post ischaemic reperfusion injury; malignancy; autoimmune disease;
KW immune system disorder; neurodegeneration; infection; gene therapy;
KW blood additive; extracorporeal circulation system.
XX
OS Vaccinia virus
XX
FH Key Location/Qualifiers
FT Misc-difference 14
FT /note= "encoded by CGA"
XX
PN WO9944625-A1.
XX
PD 10-SEP-1999.
XX
PF 02-MAR-1999; 99WO-US004635.
XX
PR 03-MAR-1998; 98US-0076821
XX
PA (UYJO) UNIV JOHNS HOPKINS.
PA (UYPI-) UNIV PITTSBURGH.
XX
PI Rosengard AM, Ahearn JM;
XX
DR WPI; 1999-550981/46.
DR N-PSDB; AAZ21090.
XX
PT New smallpox inhibitor of complement enzyme protein, used to treat
PT complement mediated disease, particularly hyperacute rejection
XX
PS Claim 1; Fig 2; 88pp; English.
XX
CC The present sequence represents Vaccinia virus smallpox inhibitor of
CC complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC activation, and so can be used to treat or prevent complement mediated
CC disorders, especially hyperacute rejection, inflammation or post
CC ischaemic reperfusion injury, malignancies, autoimmune diseases,
CC immune system disorders, neurodegeneration and infections. Hyperacute
CC rejection may also be prevented by treating the graft with SPICE before
CC transplanting it or by using a xenograft that has been transformed to
CC express SPICE from a gene therapy vector. SPICE is also useful as
CC additive to blood, e.g. in an extracorporeal circulation system (coated
CC on tubing) or in storage, also for studying complement activation.
CC Transgenic animals that express SPICE are used as sources of xenografts
XX
SQ Sequence 263 AA;

Query Match 26.3%; Score 197.5; DB 20; Length 263;
Best Local Similarity 29.9%; Pred. No. 9.5e-12;
Matches 38; Conservative 20; Mismatches 58; Indels 11; Gaps 4;

Qy 2 TSCDPPPEVKNARKPYYSLPIVIGIVLRYTCSPSYKEIGEKAIFCISENQVHATWDKAPP 61
Db 146 VKCQLPPSISNGRHNGYNDFYTIGSVVTYSCNSGYSLIGNSGVLCSGE CEWSN PP 199

Qy 62 IQESVNKTTSCSDFIVPGGEMNKGSKAPFRHGDSVIFICKANFIMKGSK IIVWCQANEMWG 121
Db 200 TCQ- - IVKCTPPETIINGYISSQFKYSYNDNVDFICKYGYKLSGSSSSICSPGNTWD 255

Qy 122 PTALPVC 128
Db 256 PTELPKC 262

RESULT 20
AAY29860
ID AAY29860 standard; Protein; 263 AA.
XX
AC AAY29860;
XX
DT 16-NOV-1999 (first entry)
XX
DE Mutated VCP giving SPICE protein sequence.
XX
KW Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW fusion protein; hyperacute rejection; xenograft; inflammation;
KW post ischaemic reperfusion injury; malignancy; autoimmune disease;
KW immune system disorder; neurodegeneration; infection; gene therapy;
KW blood additive; extracorporeal circulation system.
XX
OS Vaccinia virus
OS Synthetic.
XX
PN WO9944625-A1
XX
PD 10-SEP-1999.
XX
PF 02-MAR-1999; 99WO-US004635.
XX
PR 03-MAR-1998; 98US-0076821.

```
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Rosengard AM,  Ahearn JM;
XX
DR   WPI; 1999-550981/46.
DR   N-PSDB; AAZ21092.
XX
PT   New smallpox inhibitor of complement enzyme protein, used to treat
PT   complement-mediated disease, particularly hyperacute rejection -
XX
PS   Disclosure; Fig 3; 88pp; English
XX
CC   The present invention describes the Vaccinia virus smallpox inhibitor of
CC   complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC   activation, and so can be used to treat or prevent complement-mediated
CC   disorders, especially hyperacute rejection, inflammation or post-
CC   ischaemic reperfusion injury, malignancies, autoimmune diseases,
CC   immune system disorders, neurodegeneration and infections. Hyperacute
CC   rejection may also be prevented by treating the graft with SPICE before
CC   transplanting it or by using a xenograft that has been transformed to
CC   express SPICE from a gene therapy vector. SPICE is also useful as
CC   additive to blood, e.g. in an extracorporeal circulation system (coated
CC   on tubing) or in storage, also for studying complement activation.
CC   Transgenic animals that express SPICE are used as sources of xenografts.
CC   The present sequence represents the protein encoded by a specifically
CC   claimed mutated vaccinia complement control protein (VCP) encoding SPICE,
CC   in which position 5 of the protein is occupied by Ser (from VCP) rather
CC   than Arg.
XX
SQ   Sequence   263 AA;

  Query Match             26.4%;  Score 197.5;  DB 20;  Length 263;
  Best Local Similarity   29.9%;  Pred. No. 9.5e-12;
  Matches   48;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |:     |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSN-PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
             |:     : |  | :  |::: | |  : : |:| ||||  : : || :  |     |
Db     200 TCQ    IVKCPHPTILNGYLSSGPKRSYSYNDNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     122 PTALPVC 128
           |  || |
Db     256 P-ELPKC 261


RESULT 21
AAH48846
ID   AAH48846 standard; Protein; 263 AA.
XX
AC   AAH48846;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Cowpox virus inflammation modulatory protein (IMP).
XX
KW   Inflammation modulatory protein; IMP; cowpox virus;
KW   viral complement control protein; DAF homologue;
KW   complement activation inhibition; xenograft rejection.
XX
OS   Cowpox virus.
XX
PN   WO200071147-A1.
XX
PD   30-NOV-2000.
XX
PF   24-MAY-2000; 2000WO-US14203.
XX
PR   25-MAY-1999;  99US-0136144.
XX
PA   (KING-) KING FAISAL SPECIALIST HOSPITAL & RES CE.
PA   (JYLO-) UNIV LOUISVILLE RES FOUND INC
XX
PI   Kotwal GJ,  Al-Mohanna F,  Parhar R;
XX
DR   WPI; 2001-043865/06.
DR   N-PSDB; AAC87530.
XX
PT   Use of viral inflammation modulatory protein for blocking xenograft
PT   rejection in a patient -
XX
PS   Claim 1; Page 26; 71pp; English.
XX
CC   The invention relates to a novel method of preventing or treating
CC   xenograft rejection in a patient. The method comprises administering an
CC   inflammation modulatory protein (IMP) from cowpox virus, which acts as
CC   an inhibitor of complement activation. The invention also relates to
CC   a pharmaceutical composition comprising the IMP, suitable for the
CC   prevention or treatment of xenograft rejection. IMP is structurally
CC   related to the family of complement control proteins that includes DAF,
CC   and is functionally similar to the soluble complement receptor 1 (sCR1).
CC   IMP can block complement activation at a very early stage by binding to
CC   C3b and C4b, by inhibiting C3 convertase formation by either the
CC   classical or alternative pathway, and by accelerating the decay of the
CC   C3 convertase. Unlike other complement inhibitors, IMP retains most of
CC   the essential domains for complement control due to evolutionary
CC   pressure, and unlike sCR1, cleaves C3 at only the first cleavage site in
CC   the presence of factor 1. In addition, the viral protein is very small
CC   and is not recognised as a foreign protein because of its close
CC   structural similarity to its human homologues. The method and composition
CC   of the invention are useful for treating and preventing xenograft
CC   rejection in a patient in need of such treatment. The present sequence
CC   represents cowpox virus IMP.
XX
SQ   Sequence   263 AA;

  Query Match             26.0%;  Score 195.5;  DB 22;  Length 263;
  Best Local Similarity   30.7%;  Pred. No. 1.5e-11;
  Matches   35;  Conservative   17;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || | | |   |     |:|: |:    | ||   | |         |   ||
Db     146 VKCESPPSVTNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGLVCSG-----GEWSD-PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
             |:     : |  | :  |::: | |  :   |:|   |:  : : || :  |     |
Db     200 TCQ    IVKCPHPTISNGYLSSGEKRSYSYNDNVDFKCKHGYKLSGSSSSTCSPGNTWQ 255

Qy     122 PTALPVC 128
           |   | |
Db     256 P-ELPKC 261


RESULT 22
AAP92003
ID   AAP92003 standard; protein; 263 AA.
XX
AC   AAP92003;
XX
DT   17-DEC-2001  (updated)
DT   09-FEB-1990  (first entry)
XX
DE   Deduced sequence of complement 4b (C4b) binding protein.
XX
KW   Vaccinia virus WR strain; anti-complement protein; complement
KW   inactivators, complement 4b; C4b.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..19
FT                   /label= Signal_peptide
FT   Duplication     20
FT                   /note= "See note a in comments below."
```

```
FT   Duplication      83
FT                    /note  "See note a in comments below."
FT   Duplication      144
FT                    /note  "See note a in comments below."
FT   Duplication      202
FT                    /note  "See note a in comments below."
XX
PN   US5723208-A.
XX
PD   14 MAR 1989.
XX
PF   20 AUG 1988;    88US-234208.
XX
PR   20 AUG 1988;    88US-234208.
XX
PA   (USSH ) NAT INST OF HEALTH.
XX
PI   Kotwal G;
XX
DR   WPI; 1989-165451/22.
DR   N-PSDB; AAN90113.
XX
PT   New protein with anti complement activity
PT     encoded by Vaccinia virus 35K gene
XX
PS   Disclosure; Figure 2A; 20pp; English.
XX
CC   C4b binding protein which specifically blocks human complement cascades.
CC   It is the deduced sequence of a 35kDa protein encoded by sequence 52-840
CC   of the 35K gene of vaccinia virus strain WR. Note a - these sites
CC   indicate the start of 60 amino acid tandem repeating units which have a
CC   consensus sequence. The signal peptide sequence is not found in purified
CC   35K protein recovered from the medium of cells infected with vaccinia
CC   virus strain WR. A suggested use is to treat diseases due to abnormally
CC   high complement activity.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/ip/updates/ntis.html )
XX
SQ   Sequence   263 AA;

  Query Match             25.7%;  Score 194.5;  DB 10;  Length 263;
  Best Local Similarity   29.1%;  Pred. No. 2.4e-11;
  Matches   37;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNAKKPYYSLPIVPGIVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |    |      |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-  - GEWSD PP 199

Qy      62 ICESVNKIISCSDGIVPGGIMNKCSKAPEHMGSVIFICKANEIMKGSKIVWCQANEMWG 121
             |:      : |  | |  : |::: | |  : : |:| | ||   : : ||  :    |
Db     200 PCQ     IVKCPHPITSNGYISSGEKRSYSYNMNVDFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     122 PTALPVC 128
           |  || |
Db     256 PEELPKC 261


RESULT 23
AAY29859
ID   AAY29859 standard; Protein; 263 AA.
XX
AC   AAY29859;
XX
DT   16-NOV-1999  (first entry)
XX
DE   Vaccinia complement control protein sequence.
XX
KW   Vaccinia virus; smallpox inhibitor of complement enzyme; SPICE;
KW   fusion protein; hyperacute rejection; xenograft; inflammation;
KW   post ischaemic reperfusion injury; malignancy; autoimmune disease;
KW   immune system disorder; neurodegeneration; infection; gene therapy;
KW   blood additive; extracorporeal circulation system.
XX
OS   Vaccinia virus.
OS   Synthetic.
XX
PN   WO9944625-A1.
XX
PD   10 SEP-1999.
XX
PF   02-MAR-1999;    99WO-US04635.
XX
PR   03 MAR-1998;    98US-0076821.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Rosengard AM,  Ahearn JM;
XX
DR   WPI; 1999-550981/46.
DR   N-PSDB; AAZ21091.
XX
PT   New smallpox inhibitor of complement enzyme protein, used to treat
PT   complement-mediated disease, particularly hyperacute rejection
XX
PS   Claim 15; Fig 1; 88pp; English.
XX
CC   The present invention describes the Vaccinia virus smallpox inhibitor of
CC   complement enzymes (SPICE) protein. SPICE is an inhibitor of complement
CC   activation, and so can be used to treat or prevent complement-mediated
CC   disorders, especially hyperacute rejection, inflammation or post-
CC   ischaemic reperfusion injury, malignancies, autoimmune diseases,
CC   immune system disorders, neurodegeneration and infections. Hyperacute
CC   rejection may also be prevented by treating the graft with SPICE before
CC   transplanting it or by using a xenograft that has been transformed to
CC   express SPICE from a gene therapy vector. SPICE is also useful as
CC   additive to blood, e.g. in an extracorporeal circulation system (coated
CC   on tubing) or in storage, also for studying complement activation.
CC   Transgenic animals that express SPICE are used as sources of xenografts.
CC   The present sequence represents a vaccinia complement control protein
CC   (VCP) encoded by the specifically claimed mutated VCP nucleotide
CC   sequence, having a silent T to A transversion at nucleotide position
CC   number 267.
XX
SQ   Sequence   263 AA;

  Query Match             25.7%;  Score 194.5;  DB 20;  Length 263;
  Best Local Similarity   29.1%;  Pred. No. 2.4e-11;
  Matches   37;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNAKKPYYSLPIVPGIVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |    |      |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLCSG      GEWSD PP 199

Qy      62 ICESVNKIISCSDGIVPGGIMNKCSKAPEHMGSVIFICKANEIMKGSKIVWCQANEMWG 121
             |      : |  | |  : |::: | |  : : |:| | ||   : : ||  :    |
Db     200 PCQ     IVKCPHPITSNGYISSGEKRSYSYNMNVDFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     122 PTALPVC 128
           |  || |
Db     256 PEELPKC 261


RESULT 24
AAB13014
ID   AAB13014 standard; protein; 263 AA.
XX
AC   AAB13014;
XX
DT   11-DEC-2000  (first entry)
XX
DE   Complement inhibitory protein VCP amino acid sequence.
XX
KW   Alzheimer's disease; Vaccinia virus; VCP; complement pathway inhibitor;
```

```
KW   treatment; diagnosis; amyloid plaque.
XX
OS   Vaccinia virus.
XX
PN   WO200043027-A1.
XX
PD   27-JUL-2000.
XX
PF   19-JAN-2000; 2000WO-US001155.
XX
PR   19-JAN-1999;   99US-0116328.
XX
PA   (UYLO-) UNIV LOUISVILLE RES FOUND INC.
PA   (KOTW/) KOTWAL G J.
PA   (DALY/) DALY J.
XX
PI   Kotwal GJ, Daly J;
XX
DR   WPI; 2000-476187/41.
DR   N-PSDB; AAA72738.
XX
PT   Treating Alzheimer's disease using a Vaccinia virus protein that blocks
PT   the complementation pathway -
XX
PS   Claim 1; Page 42-43; 96pp; English.
XX
CC   This invention relates to a method for treating Alzheimer's disease. The
CC   method uses a composition comprising a Vaccinia virus complement control
CC   protein, which blocks the complement pathway by binding to complement
CC   components. The protein designated VCP, blocks complement activation and
CC   can bind to amyloid plaques in samples. The transmembrane domain and a
CC   portion of the C-terminus (Abeta) of the amyloid precursor protein (APP)
CC   form the nucleus of the amyloid plaque, the hallmark of Alzheimer's
CC   disease. The Abeta peptide activates the complement pathway. The VCP
CC   protein used in the method of the invention, down regulates the
CC   complement pathway activation caused by Abeta. The method and composition
CC   may be used for the treatment of Alzheimer's disease. The VCP protein
CC   also binds to amyloid plaques and may be labelled and used to detect the
CC   presence of amyloid plaques in pathological samples, and therefore
CC   diagnose Alzheimer's disease. The present sequence represents the
CC   Vaccinia virus VCP protein used in the method of the invention.
XX
SQ   Sequence   263 AA;

  Query Match             25.7%;  Score 193.5;  DB 21;  Length 263;
  Best Local Similarity   29.1%;  Pred. No. 2.4e-11;
  Matches   37;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLKYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 61
           : |  || : | |   |       |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQSPPSISNGRHNGYEDPYTPDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSD-PP 199

Qy      62 ICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
             |:     : |  | :  |::: | |  : : |:| | ||  : : || :  |     |
Db     200 ICQ-  --IVKCPHPTISNGYLSSGERRSYSYNDNVIFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     122 PTALPVC 128
           |  || |
Db     256 PELPKC 261


RESULT 25
AAR15233
ID   AAR15233 standard; Protein; 324 AA.
XX
AC   AAR15233;
XX
DT   17-MAR-1992  (first entry)
XX
DE   CD46 from clone pm5.8
XX
KW   SCR; short consensus repeats; transmembrane; cytoplasmic;
KW   membrane cofactor protein; MCP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..34
FT                   /label= sig_peptide
FT   Protein         34..324
FT                   /label= mat_protein
FT   Modified-site   83
FT                   /label= N-glycosylation_site
FT   Modified-site   114
FT                   /label= N-glycosylation_site
FT   Domain          289..304
FT                   /label= hydrophobic_transmembrane_domain
XX
PN   WO9118097-A.
XX
PD   28-NOV-1991.
XX
PF   10-MAY-1991;   91WO-AU00199.
XX
PR   11-MAY-1990;   90AU-0000133.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Purcell DFJ,  Russell SM,  Mckenzie JFC;
XX
DR   WPI; 1991-369251/50.
DR   N-PSDB; AAQ14919.
XX
PT   New CD46 membrane co-factor protein variants - useful as probes
PT   to identify CD46 isoforms and for diagnosing spontaneous
PT   abortion, inhibiting immuno-response and treating leukaemia
XX
PS   Disclosure; Fig 1A and 3B; 77pp; English.
XX
CC   The sequence of the pm5.8 clone is identical to the pm5.1 clone in
CC   the portion encoding the NH2 leader and four SCR regions.  The
CC   sequence after nucleotide 890 was different, and results from
CC   reading through of the cDNA into an intron sequence after the
CC   fourth SCR.  The protein encoded by this new sequence encodes
CC   a 16 amino acid hydrophobic region, probably serving as a membrane
CC   spanning sequence.  The putative cytoplasmic tail would also
CC   contain a high proportion of charged amino acids.
CC   See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
XX
SQ   Sequence   324 AA;

  Query Match             25.2%;  Score  8?.5;  DB 12;  Length 324;
  Best Local Similarity   31.4%;  Pred No. 7.7e-11;
  Matches   43;  Conservative   22;  Mismatches   51;  Indels   21;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIV----PGTVLKYTCSPSY-   YRLIGEKAIFCISENQ 91
           :: | |||:::|| :  :  : :     | |   |:|          :  |||  |:|  :|
Db     159 KVLCPPPKIKNCKHTPSEVEVFEYLDPVT-  YSCDPAPGPDPFSLIGESTIYC GDNS 214

Qy      52 VHATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKT 111
           |     :| | |    | : |  |:|  |    |    | :  :| | |   | : ||| |
Db     215 V---WSKAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDI 267

Qy     112 VWCQANEMWGPTALPVC 128
           : | :|  | |  :| |
Db     268 IVCDSNSTWDP-PVPKC 283


RESULT 26
AAR15232
ID   AAR15232 standard; Protein; 336 AA.
XX
AC   AAR15232;
XX
DT   17-MAR-1992  (first entry)
XX
```

```
DE   CD46 from clone pm5.3
XX
KW   SCR; short consensus repeats; transmembrane; cytoplasmic;
KW   membrane cofactor protein; MCP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified site   49
FT                   /label= N-glycosylation site
FT   Modified site   80
FT                   /label= N-glycosylation site
FT   Modified site   239
FT                   /label= N-glycosylation site
FT   Region          254..266
FT                   /note= "Ser/Thr rich region - o-linked
FT                   glycosylation sites"
FT   Domain          281..296
FT                   /label= hydrophobic transmembrane domain
XX
PN   WO9118097-A.
XX
PD   28-NOV-1991.
XX
PF   10-MAY-1991;   91WO-AU00199.
XX
PR   11-MAY-1990;   90AU-0000133.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Purcell DFJ, Russell SM, Mckenzie JFC;
XX
DR   WPI; 1991-369251/50.
DR   N-PSDB; AAQ14918.
XX
PT   New CD46 membrane co-factor protein variants - useful as probes
PT   to identify CD46 isoforms and for diagnosing spontaneous
PT   abortion, inhibiting immuno-response and treating leukaemia
XX
PS   Disclosure; Fig 1B and 4A; 77pp; English.
XX
CC   The pm5.3 encoding sequence also lacks the 93 bp fragment absent in
CC   pm5.10, but has an additional deletion deletion of 42 bp from a
CC   region encoding the second half (14 amino acids) of the Ser/Thr
CC   rich region removing two potential o-linked carbohydrate addition
CC   sites and five kink forming Pro residues. Since loss of this
CC   segment of DNA does not change the reading frame the remainder of
CC   the CD46 protein encoded by this clone would have the same amino
CC   acid sequence as that encoded by pm5.10.
CC   The sequence has been adjusted to corresp. to the sequence of Fig 3.
CC   See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
XX
SQ   Sequence   346 AA;

  Query Match             25.2%;  Score 189.5;  DB 12;  Length 346;
  Best Local Similarity   31.4%;  Pred. No. 8.1e-11;
  Matches   44;  Conservative   22;  Mismatches   51;  Indels   21;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIV - PGTVLRYTCSPS - - YRLIGEKAIFCISENQ 51
           :: | ||| :|| : : :       | |   |:| |:     : |||| |:|  :|
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDPVT--YSCDPAPGPDPFSLIGESTIYC GDNS 180

Qy      52 VHATWDKAPPTCESVNKTISCSDPTIVPGGFMNKGSKAIFRHGDSVTFTCKANFTMKGSKI 111
           |   | :| | |    | : | |:|  |     |   | :  :| | |   | : || |
Db     181 V--WSRAAPEC----KVVKCRFVESVFNGKQGSGCSKEYYNATVMFECDKGFYLDGSDT 233

Qy     112 VWCQANEMWGPTALPVC 128
           : | :|  | |  :| |
Db     234 IVCDSNSTWDPPVPKC 249


RESULT 2
AAR15231
```

```
ID   AAR15231 standard; Protein; 350 AA.
XX
AC   AAR15231;
XX
DT   17-MAR-1992 (first entry)
XX
DE   CD46 from clone pm5.10.
XX
KW   SCR; short consensus repeats; transmembrane; cytoplasmic;
KW   membrane cofactor protein; MCP.
XX
OS   Homo Sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   49
FT                   /label= N-glycosylation site
FT   Modified-site   80
FT                   /label= N-glycosylation site
FT   Modified site   239
FT                   /label= N-glycosylation_site
FT   Region          253..280
FT                   /note= "Ser/Thr rich region - o-linked
FT                   glycosylation sites"
FT   Domain          295..305
FT                   /label= hydrophobic transmembrane domain
XX
PN   WO9118097-A.
XX
PD   28-NOV-1991.
XX
PF   10-MAY-1991;   91WO-AU00199.
XX
PR   11-MAY-1990;   90AU-0000133.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Purcell DFJ, Russell SM, Mckenzie JFC;
XX
DR   WPI; 1991-369251/50.
DR   N-PSDB; AAQ14917.
XX
PT   New CD46 membrane co-factor protein variants - useful as probes
PT   to identify CD46 isoforms and for diagnosing spontaneous
PT   abortion - inhibiting immuno-response and treating leukaemia
XX
PS   Disclosure; Fig 1B and 4A; 77pp; English.
XX
CC   The pm5.10 coding sequence lacks 93bp of nucleotide sequence that
CC   encodes 16 of the 23 amino acids of the cytoplasmic tail and a portion
CC   of the 3' untranslated sequence of the pm5.1 clone. Since this 93bp
CC   fragment contains the termination codon ending the ORF of pm5.1,
CC   the new reading frame of pm5.10 includes a segment of sequence
CC   that forms the initial portion of the 3' untranslated sequence in
CC   pm5.1. Therefore, the CD46 mol. encoded by pm5.10 has a
CC   cytoplasmic tail of 33 amino acids, 23 of which differ from those
CC   encoded by pm5.1. This new tail would be longer by 7 amino acids,
CC   but would also have a high proportion of charged amino acids.
CC   The pm5.10 clone is not full-length and lacks the sequence at the
CC   5'end that would encode the leader peptide. The remainder of the
CC   sequence is identical to pm5.1. The pm5.10 sequence is identical
CC   to that of a MCP clone described by Lublin et al., J. Exp. Med.,
CC   168, 181-194 (1988).
CC   The sequence has been adjusted to corresp. to the sequence of Fig 3.
CC   See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
XX
SQ   Sequence   350 AA;

  Query Match             25.2%;  Score 189.5;  DB 12;  Length 350;
  Best Local Similarity   31.4%;  Pred. No. 8.3e-11;
  Matches   44;  Conservative   22;  Mismatches   51;  Indels   21;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIV - PGTVLRYTCSPS     YRLIGEKAIFCISENQ 51
           :: | |:|::|| : : : :     | |   |:| |:     : |||| |:|  :|
```

```
Db    125 KVLCTPPPKIKNGKHTFSEVEVFEYLDPVT---YSCDPAPGPDPFSLIGESTIYC-GDNS 180

Qy     52 VHATWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKT 111
          |   | : | | |     | : |  |:|  |     |     | :  :| | |    | : || |
Db    181 V---WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDT 233

Qy    112 VWCQANEMWGPTALPVC 128
          : | :|  | |  :| |
Db    234 IVCDSNSTWDP-PVPKC 249


RESULT 28
AAR10927
ID   AAR10927 standard; Protein; 357 AA.
XX
AC   AAR10927;
XX
DT   09-MAY-1991  (first entry)
XX
DE   Human membrane cofactor protein isoform with CYT1 C-terminus.
XX
KW   Mammalian membrane cofactor protein; complement activity;
KW   inflammation; autoimmune diseases; tissue injury.
XX
FH   Key             Location/Qualifiers
FT   Region          1..251
FT                   /label= short consensus repeats (4)
FT   Region          252..266
FT                   /label= serine-threonine rich region, STa
FT   Region          267..281
FT                   /label= STb
FT   Region          282..295
FT                   /label= STc
FT   Region          296..309
FT                   /label= unknown region
FT   Region          310..341
FT                   /label= hydrophobic region
FT   Region          342..357
FT                   /label= CYT1 cytoplasmic region
XX
PN   WO9102002-A.
XX
PD   21-FEB-1991.
XX
PF   20-JUL-1990;    90WO-US04107.
XX
PR   19-APR-1990;    90US-0510709.
PR   21-JUL-1989;    89US-0384210.
XX
PA   (UNIW-) UNIV OF WASHINGTON.
XX
PI   Atkinson JP;
XX
DR   WPI; 1991-073491/10.
XX
PT   New recombinant mammalian membrane co-factor protein - for
PT   treating diseases with altered complement activity e.g.
PT   inflammatory and auto-immune conditions
XX
PS   Disclosure; Fig 8; 38pp; English.
XX
CC   This is a human isoform of a membrane cofactor protein (MCP).
CC   Pharmaceutical compsns. contg. the protein are useful in the
CC   treatment of inflammatory and autoimmune diseases mediated by
CC   excess or misdirection of complement activity, e.g. rheumatoid
CC   arthritis or multiple sclerosis.  Protection against tissue
CC   injury caused by e.g. myocardial infarction or stroke may also
CC   be provided.  Antibodies raised against the MCP can be used in
CC   autoimmune disease diagnosis to predict the probability of
CC   recurrent miscarriages by testing for MCP levels in the placenta.
CC   See also AAQ10864-66.
XX
SQ   Sequence   357 AA;
```

```
  Query Match             25.2%;  Score 189.5;  DB 12;  Length 357;
  Best Local Similarity   30.2%;  Pred. No. 8.7e-11;
  Matches   42;  Conservative   23;  Mismatches   59;  Indels   15;  Gaps    6;

Qy      1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKATFCISENQVHA 54
          :: | |||::|| :  :  : :      : |:  |:     : ||||  |:|  :| |
Db    125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 181

Qy     55 TWDKAPPICESVNKTISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           | :| | |    | : |  |:|  |     |     | :  :| | |    | : |  |:|
Db    182 -WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy    115 QANEMWGPTALPVCESDFP 133
          :|  | |  :| |    |
Db    237 DSNSTWDP-PVPKCLKVLP 254


RESULT 29
AAR15230
ID   AAR15230 standard; Protein; 373 AA.
XX
AC   AAR15230;
XX
DT   17-MAR-1992  (first entry)
XX
DE   CD46 from clone pm5.6.
XX
KW   SCR; short consensus repeats; transmembrane; cytoplasmic;
KW   membrane cofactor protein; MCP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..34
FT                   /label= sig_peptide
FT   Protein         34..373
FT                   /label= mat_protein
FT   Modified-site   83
FT                   /label= N-glycosylation_site
FT   Modified-site   114
FT                   /label= N-glycosylation_site
FT   Modified-site   273
FT                   /label= N-glycosylation_site
FT   Region          287..314
FT                   /note= "Ser/Thr rich region - O-linked
FT                   glycosylation sites"
FT   Domain          329..339
FT                   /label= hydrophobic_transmembrane_domain
XX
PN   WO9118097-A
XX
PD   28-NOV-1991
XX
PF   10-MAY-1991;    91WO-AU00199.
XX
PR   11-MAY-1990;    90AU-0000133.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Purcell DFJ,  Russell SM,  Mckenzie IFC;
XX
DR   WPI; 1991-369251/50.
DR   N-PSDB; AAQ14916.
XX
PT   New CD46 membrane co-factor protein variants - useful as probes
PT   to identify CD46 isoforms and for diagnosing spontaneous
PT   abortion, inhibiting immuno-response and treating leukaemia
XX
PS   Disclosure; Fig 1B and 3A; 77pp; English.
XX
CC   The pm5.6 coding sequence corresponds to pm5.1, with the exception
CC   that as for pm5.3, a 93 bp fragment is deleted from the fourth region
```

```
CC   encoding the cytoplasmic tail.  Additionally, a 47 bp fragment is
CC   deleted from the transmembrane region (encoding the last 12 amino
CC   acids of it) due to the use of a cryptic splice acceptor signal.
CC   The absence of this fragment changes the reading frame of the
CC   C-terminal end of CD46 to encode a peptide having 34 amino acids
CC   distinct from those encoded by other CD46 nucleic acids.
CC   See also AAQ14915.25, AAQ15211.12 and AAR15457.59.
XX
SQ   Sequence    328 AA;

  Query Match             25.2%;  Score 189.5;  DB 12;  Length 328;
  Best Local Similarity   31.4%;  Pred. No. 9.2e-11;
  Matches   44;  Conservative   22;  Mismatches   51;  Indels   21;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIV    PGTVLRYTCSPS     YRLIGEKAIFCISENQ 51
           :: | |||:|| :  :  : :      | |   |:| |:     : ||||  |:|  :|
Db     159 KVLCTPPPKIKNGKHTFSEVEVEYLDPVT   YSCDPAPGPDPFSLIGESTIYC GDNS 214

Qy      52 VHATWDKAPPICESVNKTISCSDPIVPGGFMNRGSKAPFPHCDSVIFTCKANFTMKGSKT 111
           |   | :| | |    | : |  |:|  |    |  : | | |   | : || |
Db     215 V   WSRAAPEC    KVVK DRTPVVENGKQISCFGKEFYYKATVMFECDRGFYLDGSDT 267

Qy     112 VWCQANEMWGPTALPVC 128
           : | :|  | |  :| |
Db     268 IVCDSNSTWDP-PVPKC 283


RESULT 3
AAR15229
ID   AAR15229 standard; Protein; 377 AA.
XX
AC   AAR15229;
XX
DT   17-MAR-1992  (first entry)
XX
DE   CD46 from clone pm5.1.
XX
KW   SCR; short consensus repeats; transmembrane; cytoplasmic;
KW   membrane cofactor protein; MCP.
XX
OS   Homo sapiens
XX
FH   Key             Location/Qualifiers
FT   peptide         1..34
FT                   /label= sig_peptide
FT   Protein         34..377
FT                   /label= mat_protein
FT   Modified-site   83
FT                   /label= N-glycosylation_site
FT   Modified-site   114
FT                   /label= N-glycosylation_site
FT   Modified-site   273
FT                   /label= N-glycosylation_site
FT   Region          287..314
FT                   /note= "Ser/Thr rich region - 0-11kcal
FT                         glycosylation sites"
FT   Domain          329..351
FT                   /label= hydrophobic transmembrane_domain
XX
PN   WO9117267-A.
XX
PD   28-NOV-1991.
XX
PF   15-MAY-1991;   91WO-AU00199.
XX
PR   11-MAY-1990;   90AU-0000133.
XX
PA   (UYME-) UNIV MELBOURNE.
XX
PI   Purcell DFJ,  Russell SM,  McKenzie IFC;
XX
DR   WPI; 1991-369251/50.
DR   N-PSDB; AAQ14915.
```

```
XX
PT   New CD46 membrane cofactor protein variants - useful as probes
PT   to identify CD46 isoforms and for diagnosing spontaneous
PT   abortion, inhibiting immunosuppression and treating leukaemia
XX
PS   Disclosure; Fig 1A; 77pp; English.
XX
CC   The protein has four regions: (1) four contiguous homologous units
CC   of ca. 60 amino acids (SCR short consensus repeats); (2) Ser Thr
CC   rich sequence of ca. 45 amino acids; (3) transmembrane domain of
CC   ca. 23 amino acids; and (4) a 26 amino acids cytoplasmic domain
CC   contg. predominantly charged amino acids.
CC   The mature protein (38.454 D) is identical to that obtained
CC   experimentally from purified CD46 (Purcell et al., (1990)
CC   Immunogenetics, 41:21-28).
CC   See also AAQ14915.25, AAQ15211.12 and AAR15457.59.
XX
SQ   Sequence    377 AA;

  Query Match             25.2%;  Score 189.5;  DB 12;  Length 377;
  Best Local Similarity   31.4%;  Pred. No. 9.2e-11;
  Matches   44;  Conservative   22;  Mismatches   51;  Indels   21;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIV ---PGTVLRYTCSPS  -- YRLIGEKAIFCISENQ 51
           :: | |||:|| :  :  : :      | |   |:| |:     : ||||  |:|  :|
Db     159 KVLCTPPPKIKNGKHTFSEVEVEYLDPVT-- YSCDPAPGPDPFSLIGESTIYC GDNS 214

Qy      52 VHATWDKAPPICESVNKTISCSDPIVPGGFMNRGSKAPFPHCDSVIFTCKANFTMKGSKT 111
           |   | :| | |    | : |  |:|  |    |  : | | |   | : || |
Db     215 V---WSRAAPEC ---KVVKDRTPVVENGKQISCFGKEFYYKATVMFECDRGFYLDGSDT 267

Qy     112 VWCQANEMWGPTALPVC 128
           : | :|  | |  :| |
Db     268 IVCDSNSTWDP-PVPKC 283


RESULT 31
AAR41361
ID   AAR41361 standard; Protein; 251 AA.
XX
AC   AAR41361;
XX
DT   09-MAR-1994  (first entry)
XX
DE   Membrane cofactor protein fragment.
XX
KW   MCP; membrane cofactor protein; transmembrane;
KW   vascular inflammation; shock; graft rejection; SCR;
KW   intracellular.
XX
OS   Homo sapiens.
XX
PN   WO9317122-A.
XX
PD   02-SEP-1993.
XX
PF   22-FEB-1993;   93WO-JP00207.
XX
PR   21-FEB-1992;   92JP-0035445.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Kanamori T,  Nii A,  Seya T;
XX
DR   WPI; 1993-288425/36.
DR   N-PSDB; AAQ46164.
XX
PT   New protein comprising natural membrane co-factor protein lacking
PT   trans-membrane or intracellular region - for vascular
PT   inflammation, shock, graft rejection, etc.
XX
PS   Claim 3; Page 26; 81pp; Japanese.
XX
```

```
CC   The sequences (AAQ46163-64) are derived from natural MCP
CC   (see patent WO91/02002) and are lacking either the transmembrane or
CC   intracellular region. Sequence (AAQ46164) contains the first 753 bases
CC   of (AAQ46163) plus a stop codon. Both sequences contain 4 SCR regions
CC   of natural MCP
CC   Primers (AAQ46165-66) and (AAQ46169-70) are used for amplification of
CC   the DNAs encoding these proteins.
XX
SQ   Sequence   251 AA;

  Query Match             25.1%;  Score 188.5;  DB 14;  Length 251;
  Best Local Similarity   30.6%;  Pred. No. 7.2e-11;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : ||||  |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 181

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     182  WSRAAPEC   KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     237 DSNSTWDP-PVPKC 249


RESULT 32
AAW06880
ID   AAW06880 standard; Protein; 254 AA.
XX
AC   AAW06880;
XX
DT   18-MAR-1997  (first entry)
XX
DE   Membrane co-factor protein.
XX
KW   Complement inhibitor; membrane co-factor protein; MCP;
KW   decay accelerating factor; DAF; chimeric protien; glycosaminoglycan;
KW   heparin; cell lysis, sepsis; adult respiratory distress syndrome;
KW   reperfusion injury; cell damage.
XX
OS   Homo sapiens.
XX
PN   WO9634965-A2.
XX
PD   07-NOV-1996.
XX
PF   03-MAY-1996;   96WO-US06301.
XX
PR   05-MAY-1995;   95US-0435149.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Creasey AA,  Innis MA,  Zaror I;
XX
DR   WPI; 1996-506167/50.
XX
PT   Chimeric proteins for inhibiting complement-mediated cell lysis -
PT   comprise membrane co-factor protein and decay accelerating factor
PT   peptide sequences
XX
PS   Disclosure; Page 25; 84pp; English.
XX
CC   A portion (AAW06880) of the complement-inhibitor, membrane co-factor
CC   protein (MCP), is used in novel chimeric proteins of formula
CC   A-R1-B-R2-C, where A and C are peptides (AAW06875-79, AAW06883-90) able
CC   to bind glycosaminoglycans (esp. heparin) present on cell surfaces,
CC   R1 is a portion of MCP or decay accelerating factor (DAF, see also
CC   AAW06881), R2 is DAF when R1 is MCP or MCP when R1 is DAF, and B is a
CC   peptide that may have complement inhibitor activity.  The chimeric
CC   proteins (see also AAW06882) are directed to cell surfaces where they
CC   inhibit complement mediated cell lysis.  They are used to treat and
CC   prevent disease states in which complement plays a role, e.g
CC   sepsis, adult respiratory distress syndrome, reperfusion injury and
CC   tissue damage.
XX
SQ   Sequence   254 AA;

  Query Match             25.1%;  Score 188.5;  DB 17;  Length 254;
  Best Local Similarity   30.6%;  Pred. No. 7.3e-11;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : ||||  |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 181

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     182  WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     237 DSNSTWDP-PVPKC 249


RESULT 33
AAR41360
ID   AAR41360 standard; Protein; 279 AA.
XX
AC   AAR41360;
XX
DT   09-MAR-1994  (first entry)
XX
DE   Membrane cofactor protein fragment.
XX
KW   MCP; membrane cofactor protein; transmembrane;
KW   vascular inflammation; shock; graft rejection; SCR;
KW   intracellular.
XX
OS   Homo sapiens.
XX
PN   WO9317122-A.
XX
PD   02-SEP-1993.
XX
PF   22-FEB-1993;   93WO-JP00207.
XX
PR   21-FEB-1992;   92JP-0035445.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Kanamori T,  Nishi A,  Seya T;
XX
DR   WPI; 1993-288425/36.
DR   N-PSDB; AAQ46163.
XX
PT   New protein comprising natural membrane co-factor protein lacking
PT   trans-membrane or intracellular region - for vascular
PT   inflammation, shock, graft rejection, etc
XX
PS   Claim 1; Page 25, 81pp, Japanese.
XX
CC   The sequences (AAQ46163-64) are derived from natural MCP
CC   (see patent WO91/02002) and are lacking either the transmembrane or
CC   intracellular region. Sequence (AAQ46164) contains the first 753 bases
CC   of (AAQ46163) plus a stop codon. Both sequences contain 4 SCR regions
CC   of natural MCP
CC   Primers (AAQ46165-66) and (AAQ46169-70) are used for amplification of
CC   the DNAs encoding these proteins.
XX
SQ   Sequence   279 AA;

  Query Match             25.1%;  Score 188.5;  DB 14;  Length 279;
  Best Local Similarity   30.6%;  Pred. No. 8.2e-11;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;
```

```
Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS     YRLIGEKAIFCISENQVHA 54
           :  | |||: || :  :  :  :        : |:| |:      : ||||  |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVEFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 181

Qy      55 IWDKAPPICESVNKIISCSDPIVPGGPMNKCSKAIEFVPVPCEGVTELCKANEIMKGSKIVWC 114
           | :| | |    | :  |  |:  |    |    | :  :| | |   | :  || |: |
Db     182 -WSKAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :| |  :| |
Db     237 DSNSIWDP-PVPKC 249


RESULT 34
AAW69523
ID   AAW69523 standard; peptide; 279 AA.
XX
AC   AAW69523;
XX
DT   09-NOV-1998  (first entry)
XX
DE   rsMCP protein SEQ ID NO:25 from WO9829453 Claim 14.
XX
KW   drug; cell membrane directed drug; phospholipid; lipid bilayer;
KW   cell cortex; blood coagulation; inflammation; immunological disorder.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9829453 A1
XX
PD   09-JUL-1998.
XX
PF   05-JAN-1998;   98WO-JP000052.
XX
PR   27-DEC-1996;   96JP-035900.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Hasegawa I;  Moriyama S;
XX
DR   WPI; 1998-388051/33.
DR   N-PSDB; AAV40047.
XX
PT   Drugs containing peptide(s) with specific affinity to
PT   phospholipid(s)  such as phosphatidyl serine, for treatment of
PT   blood coagulation, inflammatory and immunological disorders
XX
PS   Claim 14; Page 81-83; 117pp; Japanese.
XX
CC   The present invention describes drug compositions which contain as
CC   an active component a peptide which has specific affinity to particular
CC   phospholipids (such as phosphatidyl ethanolamine or phosphatidyl
CC   serine), especially to phospholipids which constitute a lipid bilayer of
CC   cellular cortex and of which the concentration in the bilayer increases
CC   in cells which are abnormal (e.g. through injury, denaturation or
CC   activation). In particular, the peptide contains a sequence having
CC   phospholipid affinity and a structure of formula (1): (A1)a-(A2)b-(A3)c,
CC   where (A1) is one of two specific sequences (see AAW69516 and AAW69519),
CC   (A2) and (A3) are IRYLRIHPQSWVHQIALR, LRYLRIHPQSWVHQIALR (see AAW69517)
CC   or MEVIGCEAQNIY (see AAW69518); a = 0.5; b = 1.5, and c = 0.5. Preferred
CC   are the formulae A1-A2-A3, A2-A3, A2-A2-A3, A2-A2-A2-A3 or A2-A2
CC   (especially A2-A2-A3, A2-A2-A2-A3 or A2-A2). The sequence is linked to a
CC   peptide such as a blood factor, especially thrombomodulin, urinastatin
CC   or membrane cofactor protein. The drugs are used for the treatment and
CC   prevention of diseases involving blood coagulation, inflammatory and
CC   immunological disorders. The present sequence represents a peptide from
CC   the present invention
XX
SQ   Sequence   279 AA;

  Query Match             25.1%;  Score 188.5;  DB 19;  Length 279;
  Best Local Similarity   30.6%;  Pred. No. 8.1e-11;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS     YRLIGEKAIFCISENQVHA 54
           :: | |||:|| :  :  :  :        : |:| |:      : ||||  |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVEFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 181

Qy      55 IWDKAPPICESVNKIISCSDPIVPGGPMNKCSKAIEFVPVPCEGVTELCKANEIMKGSKIVWC 114
           | :| | |    | :  |  |:  |    |    | :  :| | |   | :  || |: |
Db     182 -WSKAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :| |  :| |
Db     237 DSNSIWDP-PVPKC 249


RESULT 35
AAR93939
ID   AAR93939 standard; Protein; 476 AA.
XX
AC   AAR93939;
XX
DT   20-MAY-1996  (first entry)
XX
DE   CD46 wild type
XX
KW   CD46; recombinant protein; short consensus repeat; SCR;
KW   regulator of complement activation; transgenic animal; pig;
KW   organ transplantation.
XX
OS   Homo sapiens
XX
PN   WO9606937 A1.
XX
PD   07-MAR-1996.
XX
PF   30-AUG-1995;   95WO-AU00553.
XX
PR   30-AUG-1994;   94AU-0007724.
XX
PA   (AUST-) AUSTIN RES INST.
XX
PI   Christiansen D,  Loveland B,  McKenzie IFC,  Milland J;
XX
DR   WPI; 1996-160368/16.
DR   N-PSDB; AAT17595.
XX
PT   Increasing prodn. of recombinant proteins, esp. CD46  by reducing
PT   the amt. of A and/or T in an A and/or T rich region of encoding gene
PT   exon
XX
PS   Disclosure; Page 32-33; 60pp; English.
XX
CC   A cDNA sequence (AAT17595) codes for wild type CD46 (AAR93939), a
CC   regulator of complement activation.  Expression of CD46 in
CC   eukaryotic host cells, e.g. COS-7 and CHO-K1, is facilitated
CC   by lowering the A+T content of the short consensus repeat regions
CC   of the CD46 gene (see AAT17596-99) to give products (AAR93940-43)
CC   useful e.g. as immunomodulators.  Modified CD46 may also be
CC   expressed on the organs of transgenic animals of use for organ
CC   transplantation.
XX
SQ   Sequence   476 AA;

  Query Match             25.1%;  Score 188.5;  DB 27;  Length 476;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS     YRLIGEKAIFCISENQVHA 54
           :: | |||:|| :  :  :  :        : |:| |:      : ||||  |:|  :| |
Db     158 KVLCTPPPKIKNGKHTFSEVEVEFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 214

Qy      55 IWDKAPPICESVNKIISCSDPIVPGGPMNKCSKAIEFVPVPCEGVTELCKANEIMKGSKIVWC 114
```

```
                | :| | |    | : |  |:|  |    |    | :  :| | |    | : || |: |
Db     215  WSRAAPFC    KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 269

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     270 DSNSTWDP PVPKC 282


RESULT 36
AAR93942
ID   AAR93942 standard; Protein; 377 AA.
XX
AC   AAR93942;
XX
DT   21 MAY 1996  (first entry)
XX
DE   CD46 construct subSCR2+3.
XX
KW   CD46; recombinant protein; short consensus repeat; SCR;
KW   regulator of complement activation; transgenic animal; pig;
KW   organ transplantation.
XX
OS   Synthetic.
XX
PN   WO9606937 A1.
XX
PD   07 MAR 1996.
XX
PF   30 AUG 1995;   95WO-AU00553.
XX
PR   30 AUG 1994;   94AU-0007724.
XX
PA   (AUST ) AUSTIN RES INST.
XX
PI   Christiansen D,  Loveland B,  McKenzie IFC,  Milland J;
XX
DR   WPI; 1996-160368/16.
DR   N-PSDB; AAT17598.
XX
PT   Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
PT   the amt. of A and/or T in an A and/or T rich region of encoding gene
PT   exon
XX
PS   Claim 12; Page 38 39; 60pp; English.
XX
CC   CD46 subSCR2+3 (AAR93942) is the product of a cDNA construct
CC   (AAT17598) obtd. by splice overlap extension PCR of wild-type CD46
CC   cDNA (AAT17595).  The A+T contents of AT-rich exons 3-5 of the
CC   gene, encoding the short consensus repeats (SCR) 2 and 3 of
CC   CD46, were lowered.  This resulted in improved prodn. in eukaryotic
CC   host cells, e.g. CHO-K1 and COS-7.  CD46 subSCR2+3 is used to
CC   prevent complement- or inflammation-mediated tissue damage, to
CC   improve immunity to tumours or viruses, to control fertilisation
CC   and to prevent spontaneous abortion.  Expression in transgenic
CC   animals, esp. pigs, provides organs suitable for transplantation.
XX
SQ   Sequence   377 AA;

  Query Match             25.1%;  Score 188.5;  DB 17;  Length 377;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 215

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFKHGDSVIFICKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |    | : || |: |
Db     216  WSRAAPFC    KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP PVPKC 283


RESULT 37
AAR93941
ID   AAR93941 standard; Protein; 377 AA.
XX
AC   AAR93941;
XX
DT   21-MAY-1996  (first entry)
XX
DE   CD46 construct subSCR3.
XX
KW   CD46; recombinant protein; short consensus repeat; SCR;
KW   regulator of complement activation; transgenic animal; pig;
KW   organ transplantation.
XX
OS   Synthetic.
XX
PN   WO9606937-A1.
XX
PD   07-MAR-1996.
XX
PF   30-AUG-1995;   95WO-AU00553.
XX
PR   30-AUG-1994;   94AU-0007724.
XX
PA   (AUST-) AUSTIN RES INST.
XX
PI   Christiansen D,  Loveland B,  McKenzie IFC,  Milland J;
XX
DR   WPI; 1996-160368/16.
DR   N-PSDB; AAT17597.
XX
PT   Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
PT   the amt. of A and/or T in an A and/or T rich region of encoding gene
PT   exon
XX
PS   Claim 12, Page 36 37, 60pp, English.
XX
CC   CD46 subSCR3 (AAR93941) is the product of a cDNA construct
CC   (AAT17597) obtd. by splice overlap extension PCR of wild type CD46
CC   cDNA (AAT17595).  The A+T content of AT-rich exon 5 of the CD46
CC   gene, encoding the short consensus repeat 3 (SCR3) region of
CC   CD46, is lowered.  This results in improved prodn. in eukaryotic
CC   host cells, e.g. CHO-K1 and COS-7.  CD46 subSCR3 is used to prevent
CC   complement- or inflammation-mediated tissue damage, to improve
CC   immunity to tumours or viruses, to control fertilisation and to
CC   prevent spontaneous abortion.  Expression in transgenic animals,
CC   esp. pigs, provides organs suitable for transplantation.
XX
SQ   Sequence   377 AA;

  Query Match             25.1%;  Score 188.5;  DB 17;  Length 377;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 215

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFKHGDSVIFICKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |    | : || |: |
Db     216 -WSRAAPFC---- KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 38
AAW27484
ID   AAW27484 standard; Protein; 377 AA.
XX
```

```
AC   AAW47484;
XX
DT   14-APR-1998 (first entry)
XX
DE   Human MCP
XX
KW   Human; MCP; surface; homologous complement restriction factor;
KW   HCRF; medical apparatus; medical dressing; surgical equipment;
KW   diagnostic kit; prevention; purification device; reduction;
KW   complement activation.
XX
OS   Homo sapiens.
XX
PN   WO9745086-A1.
XX
PD   04-DEC-1997
XX
PF   12-MAR-1997;   97WO-GB000684.
XX
PR   31-MAY-1996;   96GB-0022694.
PR   22-MAR-1996;   96GB-0006073.
PR   28-MAR-1996;   96GB-0006516.
XX
PA   (IMPE ) IMPERIAL COLLEGE.
XX
PI   Watkins NJ;
XX
DR   WPI; 1997-489521/45.
DR   N-PSDB; AAT90406.
XX
PT   Surfaces which prevent or reduce complement activation - having
PT   surface bound homologous complement restriction factor
XX
PS   Example 1; Fig 1A; 50pp; English.
XX
CC   The present sequence is human MCP, which was used in the
CC   development of a novel surface bound to a homologous complement
CC   restriction factor (HCRF), where the surface is not a surface to
CC   which HCRF binds in vivo. The surface can be used in medical
CC   apparatus (e.g. extra corporeal circulation systems, tubing,
CC   valves, membranes, pumps, oxygenators, catheters, cannulas, fluid
CC   reservoirs or prostheses), medical dressings, surgical equipment,
CC   diagnostic kits (e.g. kits for determining whether a patient has
CC   an abnormality which prevents the complement system from working
CC   normally or from being properly regulated) and purification
CC   devices (e.g. for purifying complement components). The surface
CC   can be used to reduce or prevent activation of complement, and to
CC   reduce morbidity due to complement activation
XX
SQ   Sequence   477 AA;

  Query Match             25.1%;  Score 188.5;  DB 18;  Length 477;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG TVLRYTCSPS     YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||  |:|  :| |
Db     160 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDPFSLIGESTIYC GDNSV- 215

Qy      55 IWTKAAPLIESVNKTISCSGPIVPGTMNRSFAPVRHTDSVIEFCKANFIMEGSKIVWC 114
             | :|  |    | :|  |:|  |    |   | :  :| | |   | :| |:|
Db     216  WSRAAPLC     KVVKCRFPVVENGKQISGFGKKFYYKAIVMFDCTGGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP PVPKC 284


RESULT 39
AAR93040
ID   AAR93040 standard; Protein; 378 AA.
XX
AC   AAR93040;
XX
DT   20-MAY-1996 (first entry)
XX
DE   CD46 construct subSCR2.
XX
KW   CD46; recombinant protein; short consensus repeat; SCR;
KW   regulator of complement activation; transgenic animal; pig;
KW   organ transplantation.
XX
OS   Synthetic.
XX
PN   WO9606937-A1.
XX
PD   07-MAR-1996.
XX
PF   30-AUG-1995;   95WO-AU00553.
XX
PR   30-AUG-1994;   94AU-0007724.
XX
PA   (AUST-) AUSTIN RES INST.
XX
PI   Christiansen D;  Loveland B,  McKenzie IFC,  Milland J;
XX
DR   WPI; 1996-160368/16.
DR   N-PSDB; AAT17596.
XX
PT   Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
PT   the amt. of A and/or T in an A and/or T rich region of encoding gene
PT   exon
XX
PS   Claim 12; Page 34-35; 60pp; English.
XX
CC   CD46 subSCR2 (AAR93940) is the product of a cDNA construct
CC   (AAT17596) obtd. by splice overlap extension PCR of wild type CD46
CC   cDNA (AAT17595). The A+T contents of AT-rich exons 3 and 4 of the
CC   gene, encoding the short consensus repeat 2 (SCR2) region of
CC   CD46, were lowered. This resulted in improved prodn. in eukaryotic
CC   host cells, e.g. CHO-K1 and COS-7. CD46 subSCR2 is used to prevent
CC   complement- or inflammation-mediated tissue damage, to improve
CC   immunity to tumours or viruses, to control fertilisation and to
CC   prevent spontaneous abortion. Expression in transgenic animals,
CC   esp. pigs, provides organs suitable for transplantation.
XX
SQ   Sequence   378 AA;

  Query Match             25.1%;  Score 188.5;  DB 17;  Length 378;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG TVLRYTCSPS     YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||  |:|  :| |
Db     160 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDPFSLIGESTIYC GDNSV- 215

Qy      55 IWTKAAPLIESVNKTISCSGPIVPGTMNRSFAPVRHTDSVIEFCKANFIMEGSKIVWC 114
             | :|  |    | :|  |:|  |    |   | :  :| | |   | :| |:|
Db     216  WSRAAPLC     KVVKCRFPVVENGKQISGFGKKFYYKAIVMFDCTGGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP PVPKC 284


RESULT 40
AAR10924
ID   AAR10924 standard; Protein; 384 AA.
XX
AC   AAR10924;
XX
DT   09-MAY-1991 (first entry)
XX
DE   Human membrane cofactor protein isoform.
XX
KW   Mammalian membrane cofactor protein; complement activity;
```

```
KW   inflammation; autoimmune diseases; tissue injury.
XX
FH   Key             Location/Qualifiers
FT   Region          1..251
FT                   /label= short consensus repeats (4)
FT   Region          252..266
FT                   /label= serine threonine rich region, STb
FT   Region          267..280
FT                   /label= STc
FT   Region          281..294
FT                   /label= unknown region
FT   Region          295..326
FT                   /label= hydrophobic region
FT   Region          327..350
FT                   /label= CYT2 cytoplasmic tail region
XX
PN   WO9102002-A.
XX
PD   21-FEB-1991.
XX
PF   20-JUL-1990;    90WO-US04107.
XX
PR   19-APR-1990;    90US-0510709.
PR   21-JUL-1989;    89US-0384210.
XX
PA   (UNIW ) UNIV OF WASHINGTON.
XX
PI   Atkinson JP;
XX
DR   WPI; 1991-073491/10.
DR   N-PSDB; AAQ10864.
XX
PT   New recombinant mammalian membrane co-factor protein for
PT   treating diseases with altered complement activity e.g.
PT   inflammatory and auto immune conditions
XX
PS   Disclosure; fig 1; 38pp; English.
XX
CC   This human isoform of a membrane cofactor protein (MCP) is useful
CC   in a pharmaceutical compsn. for the treatment of inflammatory and
CC   autoimmune diseases mediated by excess or misdirection of comple-
CC   ment activity, e.g. rheumatoid arthritis or multiple sclerosis.
CC   Protection against tissue injury caused by e.g. myocardial in-
CC   farction or stroke may also be provided. Antibodies raised
CC   against this MCP can be used in autoimmune disease diagnosis to
CC   predict the probability of recurrent miscarriages by testing for
CC   MCP levels in the placenta.
CC   See also AAQ10865-66 and AAR10927.
XX
SQ   Sequence   384 AA;

  Query Match             25.1%;  Score 188.5;  DB 12;  Length 384;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVFVFFYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV-- 215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |    | : || |: |
Db     216  WSRAAPEC---  KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|   | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 41
AAR86316
ID   AAR86316 standard; Protein; 384 AA.
XX
AC   AAR86316;
XX
DT   15-APR-1996  (first entry)
XX
DE   Human CD46.
XX
KW   Terminal complement inhibitor protein, terminal CIP; CD59; CD46;
KW   transmembrane terminal CIP; TMTCIP; Ly6 antigen; transgenic animal;
KW   organ transplantation; glycosyl-phosphatidy.inositol; GPI.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..34
FT                   /label= Sig_peptide
XX
PN   WO9523512 A1.
XX
PD   08-SEP-1995.
XX
PF   01-MAR-1995;    95WO-US02944.
XX
PR   03-MAR-1994;    94US-0205720.
XX
PA   (ALEX-) ALEXION PHARM INC.
XX
PI   Rollins S, Rother RP, Squinto SP;
XX
DR   WPI; 1995-320335/41.
DR   N-PSDB; AAT03339.
XX
PT   Terminal complement inhibitor chimeric protein and nucleic acid
PT   esp. against human complement, useful for protecting cells from
PT   complement attack e.g. in organ transplantation
XX
PS   Disclosure; Page 66-68; 85pp; English.
XX
CC   A terminal complement inhibitor protein (CIP), is used in the
CC   construction of chimeric cDNA coding for transmembrane terminal
CC   CIP, pre1. comprising amino acids 1-7 of CD59 fused to amino
CC   acids 270-350 (the transmembrane domain) of CD46 (AAR86316). Such
CC   chimeric cDNA is incorporated into a retrovirus vector and used
CC   in the breeding of transgenic animals as a means of producing
CC   transgenic organs that are protected against human complement
CC   attack upon transplantation.
XX
SQ   Sequence   384 AA;

  Query Match             25.1%;  Score 188.5;  DB 16;  Length 384;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVFVFFYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV-- 215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |    | : || |: |
Db     216  WSRAAPEC---  KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|   | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 42
AAB58394
ID   AAB58394 standard; Protein; 421 AA
XX
AC   AAB58394;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 732.
```

ATOM	3737	C	LEU	C	45	62.664	85.961	-4.263	1.00	28.71	C
ATOM	3738	O	LEU	C	45	63.094	86.894	-3.558	1.00	21.68	C
ATOM	3739	N	CYS	C	46	61.410	85.547	-4.238	1.00	25.91	C
ATOM	3740	CA	CYS	C	46	60.455	86.152	-3.343	1.00	26.90	C
ATOM	3741	C	CYS	C	46	60.610	85.392	-2.032	1.00	30.25	C
ATOM	3742	O	CYS	C	46	60.448	84.173	-1.980	1.00	28.51	C
ATOM	3743	CB	CYS	C	46	59.021	86.026	-3.908	1.00	28.02	C
ATOM	3744	SG	CYS	C	46	57.741	86.546	-2.696	1.00	28.63	C
ATOM	3745	N	ILE	C	47	60.925	86.115	-0.971	1.00	21.11	C
ATOM	3746	CA	ILE	C	47	61.120	85.484	0.314	1.00	24.65	C
ATOM	3747	CD	ILE	C	47	62.619	85.581	0.722	1.00	32.62	C
ATOM	3748	CG2	ILE	C	47	63.505	84.878	-0.358	1.00	27.06	C
ATOM	3749	CG1	ILE	C	47	63.050	86.958	0.959	1.00	25.64	C
ATOM	3750	CD1	ILE	C	47	64.429	87.119	1.592	1.00	32.03	C
ATOM	3751	C	ILE	C	47	60.431	86.274	1.421	1.00	31.05	C
ATOM	3752	O	ILE	C	47	59.931	87.399	1.226	1.00	22.58	C
ATOM	3753	N	THR	C	48	60.421	85.669	2.596	1.00	27.58	C
ATOM	3754	CA	THR	C	48	59.964	86.388	3.764	1.00	27.51	C
ATOM	3755	CB	THR	C	48	58.584	85.987	4.273	1.00	28.85	C
ATOM	3756	OG1	THR	C	48	58.565	86.778	5.456	1.00	25.10	C
ATOM	3757	CG2	THR	C	48	58.490	84.526	4.619	1.00	26.50	C
ATOM	3758	C	THR	C	48	61.001	86.164	4.839	1.00	29.22	C
ATOM	3759	O	THR	C	48	61.455	85.049	5.090	1.00	25.46	C
ATOM	3760	N	LYS	C	49	61.445	87.243	5.463	1.00	27.29	C
ATOM	3761	CA	LYS	C	49	62.425	87.112	6.522	1.00	31.22	C
ATOM	3762	CB	LYS	C	49	63.478	88.220	6.427	1.00	26.53	C
ATOM	3763	CG	LYS	C	49	64.229	88.239	5.134	1.00	35.41	C
ATOM	3764	CD	LYS	C	49	65.703	88.351	5.346	1.00	40.12	C
ATOM	3765	CE	LYS	C	49	66.262	87.134	6.055	1.00	42.32	C
ATOM	3766	NZ	LYS	C	49	67.729	87.262	6.166	1.00	46.64	C
ATOM	3767	C	LYS	C	49	61.745	87.219	7.877	1.00	33.53	C
ATOM	3768	O	LYS	C	49	62.249	86.692	8.856	1.00	30.36	C
ATOM	3769	N	ASP	C	50	60.596	87.896	7.952	1.00	25.34	C
ATOM	3770	CA	ASP	C	50	59.978	88.083	9.253	1.00	26.98	C
ATOM	3771	CB	ASP	C	50	59.961	89.579	9.597	1.00	25.16	C
ATOM	3772	CG	ASP	C	50	59.154	90.397	8.607	1.00	27.41	C
ATOM	3773	OD1	ASP	C	50	58.459	89.768	7.749	1.00	22.49	C
ATOM	3774	OD2	ASP	C	50	59.207	91.664	8.693	1.00	26.37	C
ATOM	3775	C	ASP	C	50	58.598	87.492	9.365	1.00	25.28	C
ATOM	3776	O	ASP	C	50	57.883	87.770	10.321	1.00	24.69	C
ATOM	3777	N	LYS	C	51	58.247	86.664	8.386	1.00	26.19	C
ATOM	3778	CA	LYS	C	51	56.948	86.006	8.319	1.00	23.97	C
ATOM	3779	CB	LYS	C	51	56.717	85.091	9.526	1.00	31.09	C
ATOM	3780	CG	LYS	C	51	57.749	84.021	9.749	1.00	32.84	C
ATOM	3781	CD	LYS	C	51	57.625	82.919	8.769	1.00	44.41	C
ATOM	3782	CE	LYS	C	51	58.505	81.752	9.179	1.00	52.55	C
ATOM	3783	NZ	LYS	C	51	58.266	81.397	10.605	1.00	51.40	C
ATOM	3784	C	LYS	C	51	55.791	86.996	8.241	1.00	25.78	C
ATOM	3785	O	LYS	C	51	54.660	86.628	8.510	1.00	25.03	C
ATOM	3786	N	VAL	C	52	56.048	88.252	7.896	1.00	25.78	C
ATOM	3787	CA	VAL	C	52	54.937	89.195	7.758	1.00	23.31	C
ATOM	3788	CB	VAL	C	52	54.976	90.286	8.894	1.00	30.82	C
ATOM	3789	CG1	VAL	C	52	53.786	91.197	8.802	1.00	27.39	C
ATOM	3790	CG2	VAL	C	52	54.946	89.611	10.264	1.00	33.79	C
ATOM	3791	C	VAL	C	52	55.085	89.832	6.376	1.00	29.16	C

```
ID   AAW06882 standard; Protein; 577 AA.
XX
AC   AAW06882;
XX
DT   18-MAR-1997  (first entry)
XX
DE   Membrane co-factor protein-decay accelerating factor hybrid.
XX
KW   Complement inhibitor; membrane co-factor protein; MCP;
KW   decay accelerating factor; DAF; chimeric protien; glycosaminoglycan;
KW   heparin; cell lysis; sepsis; adult respiratory distress syndrome;
KW   reperfusion injury; cell damage.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1..254
FT                   /label= MCP
FT   Region          255..577
FT                   /label= DAF
XX
PN   WO9634965-A2.
XX
PD   07-NOV-1996.
XX
PF   04-MAY-1996;   96WO-US06301.
XX
PR   05-MAY-1995;   95US-0435149.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Creasey AA,  Innis MA,  Zaror I;
XX
DR   WPI; 1996-506167/50.
DR   N-PSDB; AAT46065.
XX
PT   Chimeric proteins for inhibiting complement-mediated cell lysis -
PT   comprise membrane co-factor protein and decay accelerating factor
PT   peptide sequences
XX
PS   Disclosure; Page 25-26; 34pp; English.
XX
CC   A hybrid protein (AAW06882) comprises portions (see also AAW06880-81)
CC   of the complement-inhibitors membrane co-factor protein (MCP) and
CC   decay accelerating factor (DAF).  It can be used in novel chimeric
CC   proteins also incorporating a peptide (AAW06875-79, AAW06883-90) able
CC   to bind glycosaminoglycans (esp. heparin) present on cell surfaces.
CC   The constructs are encoded by overlapping PCR (see also AAT46066-72)
CC   using MCP-DAF hybrid DNA (AAT46065) as template, and can be expressed
CC   in e.g. insect cells.  The chimeric proteins are directed to cell
CC   surfaces where they inhibit complement-mediated cell lysis.  They
CC   are used to treat and prevent disease states in which complement
CC   plays a role, e.g. sepsis, adult respiratory distress syndrome,
CC   reperfusion injury and tissue damage.
XX
SQ   Sequence   577 AA;

  Query Match             25.1%;  Score 188.5;  DB 17;  Length 577;
  Best Local Similarity   30.6%;  Pred. No. 2e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 181

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     182  WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     237 DSNSTWDPPVPKC 249
```

```
RESULT 45
AAE12569
ID   AAE12569 standard; Protein; 611 AA.
XX
AC   AAE12569;
XX
DT   03-JAN-2002  (first entry)
XX
DE   CAB2 protein.
XX
KW   Expression vector; crippled selectable marker; neomycin resistance;
KW   HIV protein; human immunodeficiency virus; improved expression; CAB2;
KW   CAB4; amplifiable selectable marker; dihydrofolate reductase; dhfr;
KW   transgene; continuous cell line preparation.
XX
OS   Unidentified.
XX
PN   US2001024807-A1.
XX
PD   27-SEP-2001
XX
PF   22-DEC-2000; 2000US-0748061.
XX
PR   01-NOV-1999;   99US-162930P.
PR   30-DEC-1999;   99US-0475460.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Innis M,  Scott EM;
XX
DR   WPI; 2001-638503/73.
DR   N-PSDB; AAD20355.
XX
PT   New expression vector, useful for improving expression of transgene or
PT   polypeptide, comprises 3 polynucleotides encoding crippled selectable
PT   marker, heterologous polypeptide or second amplifiable selectable
PT   marker -
XX
PS   Example 2; Fig 1; 27pp; English.
XX
CC   The invention relates to a new expression vector comprising a first
CC   polynucleotide encoding a first, crippled selectable marker which
CC   include sequences encoding antibiotic (neomycin) resistance containing
CC   one or more crippling mutations; second polynucleotide encoding a
CC   heterologous polypeptide of interest which is viral protein (e.g., an
CC   HIV protein) or is CAB2 or CAB4; and a third polynucleotide encoding
CC   a second amplifiable selectable marker e.g. dihydrofolate reductase
CC   (dhfr). The expression vectors are useful for the efficient expression
CC   of desired polypeptides or improving expression of a transgene of
CC   interest. The transformed cells can be used in the preparation of
CC   continuous cell lines in which the cells are essentially immortal or
CC   for the preparation of established cell lines that have the potential
CC   to be subcultured in vitro. The present sequence is CAB2 protein.
CC   Aberrant splicing of CAB2 DNA was corrected by removing donor and
CC   acceptor sites using overlapping PCR.
XX
SQ   Sequence   611 AA;

  Query Match             25.1%;  Score 188.5;  DB 22;  Length 611;
  Best Local Similarity   30.6%;  Pred. No. 2.1e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | | |::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :|  |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216 -WSRAAPEC----KVVKCRFPVVENGKQISGIGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
```

OM [illegible] USNSTWDP PVPRT 283

Search completed: November 6, 2002, 16:06:31
Job time : 30.1004 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 10.539 Seconds
(without alignments)
312.880 Million cell updates/sec

Title: US-09-834-309-6
Perfect score: 752
Sequence: 1 EISCDPPPEVKNAPKPYYSI ANEMWGPTALPVCESDFPLE 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	220.5	29.3	1466	6	5256642 6	Patent No. 5256642
2	220.5	29.3	1466	6	5472939 6	Patent No. 5472939
3	220.5	29.3	1537	6	5256642 9	Patent No. 5256642
4	220.5	29.3	1537	6	5472939 9	Patent No. 5472939
5	220.5	29.3	1847	6	5256642 10	Patent No. 5256642
6	220.5	29.3	1847	6	5472939 10	Patent No. 5472939
7	220.5	29.3	2039	6	5256642 2	Patent No. 5256642
8	220.5	29.3	2039	6	5472939 2	Patent No. 5472939
9	199.5	26.5	463	4	US-08-981-234B-2	Sequence 2, Appli
10	192.5	25.6	263	1	US-07-906-983-2	Sequence 2, Appli
11	190.5	25.3	384	4	US-08-139-195-2	Sequence 2, Appli
12	188.5	25.1	169	1	US-08-310-416A-18	Sequence 18, Appl
13	188.5	25.1	169	2	US-08-888-171-18	Sequence 18, Appl
14	188.5	25.1	254	1	US-08-310-416A-13	Sequence 13, Appl
15	188.5	25.1	254	2	US-08-888-171-13	Sequence 13, Appl
16	188.5	25.1	254	2	US-08-435-149-1	Sequence 1, Appli
17	188.5	25.1	293	1	US-08-310-416A-16	Sequence 16, Appl
18	188.5	25.1	293	2	US-08-888-171-16	Sequence 16, Appl
19	188.5	25.1	424	2	US-08-528-057-46	Sequence 46, Appl
20	188.5	25.1	470	2	US-08-528-057-42	Sequence 42, Appl
21	188.5	25.1	473	2	US-08-528-057-44	Sequence 44, Appl
22	188.5	25.1	477	2	US-08-528-057-2	Sequence 2, Appli
23	188.5	25.1	484	6	5514787-2	Patent No. 5514787
24	188.5	25.1	577	2	US-08-435-149-3	Sequence 3, Appli
25	188.5	25.1	611	4	US-09-475-460A-32	Sequence 32, Appl
26	178.5	23.7	133	2	US-08-356-361-31	Sequence 31, Appl
27	178.5	23.7	133	2	US-08-769-967A-31	Sequence 31, Appl
28	178.5	23.7	254	2	US-08-356-361-29	Sequence 29, Appl
29	178.5	23.7	254	2	US-08-769-967A-29	Sequence 29, Appl
30	175.5	23.3	254	2	US-08-356-361-30	Sequence 30, Appl
31	175.5	23.3	254	2	US-08-769-967A-30	Sequence 30, Appl
32	163	21.7	86	6	5514582-41	Patent No. 5514582
33	153.5	20.4	124	6	5514582-38	Patent No. 5514582
34	153.5	20.4	265	2	US-08-177-109A-57	Sequence 57, Appl
35	153.5	20.4	265	2	US-08-687-706-57	Sequence 57, Appl
36	153.5	20.4	764	2	US-08-177-109A-2	Sequence 2, Appli
37	153.5	20.4	764	2	US-08-687-706-2	Sequence 2, Appli
38	150.5	20.0	126	6	5514582-35	Patent No. 5514582
39	145.5	19.3	126	6	5514582-43	Patent No. 5514582
40	144	19.1	197	2	US-08-356-361-27	Sequence 27, Appl
41	144	19.1	197	2	US-08-769-967A-27	Sequence 27, Appl
42	142	18.9	323	2	US-08-435-149-2	Sequence 2, Appli
43	142	18.9	324	1	US-08-310-416A-14	Sequence 14, Appl
44	142	18.9	324	2	US-08-888-171-14	Sequence 14, Appl
45	139	18.5	610	1	US-08-365-470-3	Sequence 3, Appli
46	139	18.5	610	3	US-09-209-668-19	Sequence 19, Appl
47	139	18.5	610	4	US-09-009-490A-89	Sequence 89, Appl
48	139	18.5	610	6	5217870-2	Patent No. 5217870
49	130.5	17.4	274	2	US-08-177-109A-58	Sequence 58, Appl
50	130.5	17.4	274	2	US-08-687-706-58	Sequence 58, Appl
51	127.5	17.0	1019	1	US-08-296-014A-4	Sequence 4, Appli
52	127.5	17.0	1019	2	US-08-596-405-4	Sequence 4, Appli
53	127.5	17.0	1019	2	US-08-877-620-4	Sequence 4, Appli
54	127.5	17.0	1083	1	US-08-296-014A-2	Sequence 2, Appli
55	127.5	17.0	1083	2	US-08-596-405-2	Sequence 2, Appli
56	127.5	17.0	1083	2	US-08-877-620-2	Sequence 2, Appli
57	126.5	16.8	120	6	5514582-36	Patent No. 5514582
58	125.5	16.7	196	3	US-08-824-692-32	Sequence 32, Appl
59	125.5	16.7	229	3	US-08-824-692-31	Sequence 31, Appl
60	125.5	16.7	290	3	US-08-824-692-29	Sequence 29, Appl
61	124	16.5	830	6	5378464-2	Patent No. 5378464
62	123.5	16.4	484	2	US-08-252-193C-9	Sequence 9, Appli
63	123.5	16.4	484	3	US-09-276-197-9	Sequence 9, Appli
64	123.5	16.4	830	5	PCT-US91-05059-2	Sequence 2, Appli
65	120	16.0	207	2	US-08-640-977-5	Sequence 5, Appli
66	120	16.0	248	2	US-08-640-977-2	Sequence 2, Appli
67	120	16.0	266	2	US-08-640-977-4	Sequence 4, Appli
68	120	16.0	326	2	US-08-640-977-1	Sequence 1, Appli
69	119	15.8	830	1	US-08-110-158-4	Sequence 4, Appli
70	115.5	15.4	574	6	5378464-3	Patent No. 5378464
71	113	15.0	127	6	5514582-37	Patent No. 5514582
72	110.5	14.7	177	3	US-08-824-692-36	Sequence 36, Appl
73	110	14.6	181	2	US-08-640-977-3	Sequence 3, Appli
74	106	14.1	128	6	5514582-42	Patent No. 5514582
75	104.5	13.9	125	6	5514582-39	Patent No. 5514582
76	103	13.7	128	6	5514582-35	Patent No. 5514582
77	100.5	13.4	127	6	5514582-32	Patent No. 5514582
78	100.5	13.4	177	3	US-08-824-692-30	Sequence 30, Appl
79	98	13.0	62	1	US-08-210-266A-13	Sequence 13, Appl
80	98	13.0	62	1	US-08-688-675-13	Sequence 13, Appl
81	98	13.0	62	3	US-08-477-860C-13	Sequence 13, Appl
82	96	12.8	240	3	US-08-824-692-23	Sequence 23, Appl
83	95	12.6	216	3	US-08-824-692-24	Sequence 24, Appl
84	94	12.5	84	4	US-08-793-418-6	Sequence 6, Appli
85	93.5	12.4	933	2	US-08-313-200-1	Sequence 1, Appli
86	93.5	12.4	933	5	PCT-US93-03857-1	Sequence 1, Appli
87	93	12.4	62	1	US-08-210-266A-12	Sequence 12, Appl
88	93	12.4	62	1	US-08-688-675-12	Sequence 12, Appl
89	93	12.4	62	3	US-08-477-860C-12	Sequence 12, Appl
90	90.5	12.0	76	2	US-08-356-361-28	Sequence 28, Appl
91	90.5	12.0	76	2	US-08-769-967A-28	Sequence 28, Appl
92	90	12.0	17	4	US-08-602-995A-331	Sequence 331, App
93	88.5	11.8	385	1	US-08-340-540A-2	Sequence 2, Appli
94	88.5	11.8	385	2	US-08-461-592B-2	Sequence 2, Appli
95	87.5	11.6	372	2	US-08-513-278-2	Sequence 2, Appli
96	87.5	11.6	372	6	5514582-2	Patent No. 5514582
97	84.5	11.2	123	3	US-08-824-692-37	Sequence 37, Appl
98	84.5	11.2	123	3	US-08-824-692-38	Sequence 38, Appl
99	83	11.0	145	2	US-08-640-977-6	Sequence 6, Appli
100	82.5	11.0	128	6	5514582-34	Patent No. 5514582

ALIGNMENTS

RESULT 1
5256642-4
;Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLITIC AGENT, AND THE METHODS OF
;USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01-APR-1988
;SEQ ID NO:6;
; LENGTH: 1466
5256642-6

  Query Match 29.8%; Score 220.5; DB 6; Length 1466;
  Best Local Similarity 36.0%; Pred. No. 1.8e-15;
  Matches 49; Conservative 18; Mismatches 94; Indels 15; Gaps 5;

Qy 2 ISCDPPPEVKNARKPYYS-LPIVPGTVLRYTCSPS YRLIGEKAIFCISENQV 52
|||:||| : | :|| ||| | | : |:|||:|:| |::
Db 1015 ISCEPPPTISNG-DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy 53 HATWDKAPPICESVNKTISCSDPTVPGGFMNKGSKAPFFHGDSVTFTCKANFIMKGSKTV 112
| || | | || |:| | |:::| : | | | | | || ||
Db 1073 VGVWSSPPPRCISTNK-CTAPEVENAIRVPGNRSFTSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 113 WCQANEMWGPTALPVC 128
|| : ||| || |
Db 1130 QCQTNGRWGP-KLPHC 1144

RESULT 2
5472939-6
;Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01-APR-1988
;SEQ ID NO:6;
; LENGTH: 1466
5472939-6

  Query Match 29.8%; Score 220.5; DB 6; Length 1466;
  Best Local Similarity 36.0%; Pred. No. 1.8e-15;
  Matches 49; Conservative 18; Mismatches 94; Indels 15; Gaps 5;

Qy 2 ISCDPPPEVKNARKPYYS-LPIVPGTVLRYTCSPS YRLIGEKAIFCISENQV 52
|||:||| : | :|| ||| | | : |:|||:|:| |::
Db 1015 ISCEPPPTISNG-DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy 53 HATWDKAPPICESVNKTISCSDPTVPGGFMNKGSKAPFFHGDSVTFTCKANFIMKGSKTV 112
| || | | || |:| | |:::| : | | | | | || ||
Db 1073 VGVWSSPPPRCISTNK-CTAPEVENAIRVPGNRSFTSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 113 WCQANEMWGPTALPVC 128
|| : ||| || |
Db 1130 QCQTNGRWGP-KLPHC 1144

RESULT 3
5256642-5
;Patent No. 5256642
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLITIC AGENT, AND THE METHODS OF
;USE THEREOF
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,128
; FILING DATE: 24-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989
; APPLICATION NUMBER: 332,865
; FILING DATE: 03-APR-1989
; APPLICATION NUMBER: 176,542
; FILING DATE: 01-APR-1988
;SEQ ID NO:5;
; LENGTH: 1547
5256642-5

  Query Match 29.8%; Score 220.5; DB 6; Length 1547;
  Best Local Similarity 36.0%; Pred. No. 1.9e-15;
  Matches 49; Conservative 18; Mismatches 94; Indels 15; Gaps 5;

Qy 2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS -- YRLIGEKAIFCISENQV 52
|||:||| : | :|| ||| | | : |:|||:|:| |::
Db 1015 ISCEPPPTISNG-DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy 53 HATWDKAPPICESVNKTISCSDPTVPGGFMNKGSKAPFFHGDSVTFTCKANFIMKGSKTV 112
| || | | || |:| | |:::| : | | | | | || ||
Db 1073 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFTSLTEIIRFRCQPGFVMVGSHTV 1129

Qy 113 WCQANEMWGPTALPVC 128
|| : ||| || |
Db 1130 QCQTNGRWGP-KLPHC 1144

RESULT 4
5472939-5
;Patent No. 5472939
; APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
; TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
; NUMBER OF SEQUENCES: 30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,825
; FILING DATE: 19-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 588,128
; FILING DATE: 24-SEP-1990
; APPLICATION NUMBER: 412,745
; FILING DATE: 26-SEP-1989

```
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR 1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR 1988
;SEQ ID NO:5:
;       LENGTH: 1537
5472939 5

  Query Match             29.3%;  Score 220.5;  DB 6;  Length 1537;
  Best Local Similarity   36.0%;  Pred. No. 1.9e-15;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS   LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        ||| | |         : |:|||:|:| |::
Db    1015 ISCFPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1072

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1073 VGVWSSPPPRCISTNK - CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1129

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| || |
Db    1130 QCQTNGRWGP-KLPHC 1144


RESULT 5
5256642 10
;Patent No. 5256642
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLTIC AGENT, AND THE METHODS OF
;USE THEREOF
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/588,128
;       FILING DATE: 24-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:10:
;       LENGTH: 1847
5256642 10

  Query Match             29.3%;  Score 220.5;  DB 6;  Length 1847;
  Best Local Similarity   36.0%;  Pred. No. 2.5e-15;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS- -LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        ||| | |         : |:|||:|:| |::
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : : | |:  | | || ||
Db    1580 VGVWSSPPPRCISTNK   CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| || |
Db    1637 QCQTNGRWGP-KLPHC 1651


RESULT 6
5472939 10
;Patent No. 5472939
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/138,825
;       FILING DATE: 19-OCT-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 588,128
;       FILING DATE: 24-SEP-1990
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:10:
;       LENGTH: 2006
5472939-10

  Query Match             29.3%;  Score 220.5;  DB 6;  Length 1847;
  Best Local Similarity   36.0%;  Pred. No. 2.5e-15;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS---  -YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: |              : |:|||:|:| |::
Db    1522 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1579

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       | :  |   : : | |:  | | || ||
Db    1580 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1636

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| || |
Db    1637 QCQTNGRWGP-KLPHC 1651


RESULT 7
5256642 2
;Patent No. 5256642
;     APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C. JR.
;     TITLE OF INVENTION: COMPOSITIONS OF SOLUBLE COMPLEMENT
;RECEPTOR 1 (CR1) AND A THROMBOLTIC AGENT, AND THE METHODS OF
;USE THEREOF
;     NUMBER OF SEQUENCES: 30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/588,128
;       FILING DATE: 24-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 412,745
;       FILING DATE: 26-SEP-1989
;       APPLICATION NUMBER: 332,865
;       FILING DATE: 03-APR-1989
;       APPLICATION NUMBER: 176,532
;       FILING DATE: 01-APR-1988
;SEQ ID NO:2:
;       LENGTH: 2039
5256642-2

  Query Match             29.3%;  Score 220.5;  DB 6;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 2.8e-15;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS----  -YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: |              : |:|||:|:| |::
Db    1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       | :  |   : : | |:  | | || ||
Db    1575 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||| || |
```

```
Db    1632 QCQINGRWGP KLPBC 1646


RESULT 8
S47299-2
;Patent No. S47299
;  APPLICANT: FEARON, DOUGLAS T.;KLICKSTEIN, LLOYD B.;WONG,
;WINNIE W.;CARSON, GERALD R.;CONCINO, MICHAEL F.;IP, STEPHEN
;H.;MAKRIDES, SAVVAS;MARSH, HENRY C., JR.
;  TITLE OF INVENTION: METHOD OF TREATING COMPLEMENT
;MEDIATED DISORDERS
;  NUMBER OF SEQUENCES: 30
;  CORRESPONDENCE ADDRESS:
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/138,825
;    FILING DATE: 19-OCT-1993
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: 588,128
;    FILING DATE: 24-SEP-1990
;    APPLICATION NUMBER: 412,745
;    FILING DATE: 26-SEP-1989
;    APPLICATION NUMBER: 332,865
;    FILING DATE: 03-APR-1989
;    APPLICATION NUMBER: 176,542
;    FILING DATE: 01-APR-1988
;SEQ ID NO: 2:
;    LENGTH: 2039
S47299-2

  Query Match             29.3%;  Score 220.5;  DB 6;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 2.8e-15;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKEYYS   LPIVPGIVLRYICSPS----- YRLIGEKAIFCISENQV 52
           |||:||| : |    :||         ||| | |          : |:||::|:| |::
Db    1517 ISCEPPPFISNG  DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 HATWDKAPPICESVNKTISCSDIEVPGGFMNKGSKAPFRHGDSVTFICKANFIMKGSKTV 112
             |   :| | | ||    | | |       |::: |   : : | |:  | | || ||
Db    1575 VGVWSSPPPRCISTNK   CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1631

Qy     113 WCQANEMWGPTALPVC 128
           || |  :|| || |
Db    1632 QCQINGRWGP KLPBC 1646


RESULT 9
US-08-981-234B-2
; Sequence 2, Application US/08981234B
; Patent No. 6270997
;  GENERAL INFORMATION:
;    APPLICANT:  TOYOMURA, KOJI
;    APPLICANT:  MURAKAMI, HIROSHI
;    APPLICANT:  SHIGEHISA, TAMOTSU
;    TITLE OF INVENTION:  DNA ENCODING A PORCINE COMPLEMENT
;    TITLE OF INVENTION:  INHIBITOR
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  PO BOX 747
;      CITY:  FALLS CHURCH
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,234B
;      FILING DATE:  12-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURPHY JR, GERALD M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  2520-111P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  463 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-981-234B-2

  Query Match             26.5%;  Score 199.5;  DB 4;  Length 463;
  Best Local Similarity   31.9%;  Pred. No. 5.9e-14;
  Matches   45;  Conservative   21;  Mismatches   52;  Indels   23;  Gaps    6;

Qy       1 EISCDPPPEVKNAR-----KPYYSLPIVPGTVLRYICSPS ---  YRLIGEKAIFCISEN 50
           :| | || |: | :     |  :          |: |:|  |     : |:|| ::||| ::
Db     168 KILCKPPGEIPNGKYTNSHKDVFEY----NEVVTYSCLSSTGPDEFSLVGESSLFCTKD 223

Qy      51 QVHATWDKAPPICFSVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSK 110
           :    |   || |    | : |  |:|| | :  |  : | :    | | | | || |
Db     224 E----WSSPPPEC----KVVKCPYPVVPNGEIVSGFGSKFFYRAGVVKFSCNAGFTLRGPD 275

Qy     111 TVWCQANEMWGPTALPVCESD 131
           |: | ||:  | |  :| |  |
Db     276 TIVCGANSTWEP-EMPQCIKD 295


RESULT 10
US-07-906-983-2
; Sequence 2, Application US/07906983
; Patent No. 5187268
;  GENERAL INFORMATION:
;    APPLICANT:  Kotwal, Girish
;    APPLICANT:  Moss, Bernard
;    TITLE OF INVENTION:  Synthetic, Anti-Complement Protein and
;    TITLE OF INVENTION:  the Gene Encoding Same
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/906,983
;      FILING DATE:  19920701
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Weber, Kenneth A.
;      REGISTRATION NUMBER:  31,677
;      REFERENCE/DOCKET NUMBER:  15280-9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  263 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-906-983-2

  Query Match             25.6%;  Score 192.5;  DB 1;  Length 263;
```

```
  Best Local Similarity   29.1%;  Pred. No. 2.2e-13;
  Matches   47;  Conservative   13;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ESCDPPPEVKNARKPYYSLPIVPGTVLKYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 61
           : |  || : | |   |       |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQSPPSISNGRHNGYEDFYIDGSVVIYSCNSGYSLIGNSGVLCSG-----GEWSD-PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPPFHGDSVIFTCKANFTMKGSKIVWCQANEMWG 121
             |:     : |  | :  |::: | |  : : |:| | ||  : : || :  |     |
Db     200 TCQ     IVKCPHPITSNGYLSSGFKRSYSYNDNVIFKCKYCYKVSGSSSSTCSPGNTWK 255

Qy     122 PTALPVC 128
           |  || |
Db     256 P ELPKC 261


RESULT 11
US-08-139-195-2
; Sequence 2, Application US/08139195
; Patent No. 6218520
;  GENERAL INFORMATION:
;    APPLICANT:  Atkinson, John P.
;    TITLE OF INVENTION:  RECOMBINANTLY PRODUCED HUMAN MEMBRANE
;    TITLE OF INVENTION:  COFACTOR PROTEIN (MCP)
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  1100 Peachtree Street, Suite 2800
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-4530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/139,195
;      FILING DATE:  20-OCT-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/948,450
;      FILING DATE:  21-SEP-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/384,210
;      FILING DATE:  21-JUL-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU100C0N
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404)-815-6508
;      TELEFAX:  (404) 815-6555
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Human Membrane CoFactor Protein (MCP)
US-08-139-195-2

  Query Match             25.3%;  Score 190.5;  DB 4;  Length 384;
  Best Local Similarity   30.6%;  Pred. No. 6.1e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
             :: | ||| :|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVIYS DPA-GPDPGSLIGESTIYC GDNSV- 215

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPPFHGDSVIFTCKANFTMKGSKIVWC 114
             | :| | |     | : |  |:|  |    |    |  :  :| | |   | : || | |
Db     216 -WSRAAPEC----KVVKCRGPVVENGKQISGFGKRFYYKATVMEECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 12
US-08-310-416A-18
; Sequence 18, Application US/08310416A
; Patent No. 5679546
;  GENERAL INFORMATION:
;    APPLICANT:  Jone-Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION   COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 MB
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,416A
;      FILING DATE:  22-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul W. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  06180/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  169 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-310-416A-18

  Query Match             25.1%;  Score 188.5;  DB 1;  Length 169;
  Best Local Similarity   30.6%;  Pred. No. 3.5e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YPLIGEKAIFCISENQVHA 54
             :: | ||| :|| :  :  : :       : |:| |:      : |||  |:|  :| |
Db       1 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAE PDPFSLIGESTIYC-GDNSV- 57

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVIFTCKANFTMKGSKIVWC 114
             | :| | |     | : |  |:|  |    |    |  :  :| | |   | : || | |
Db      58 -WSRAAPEC----KVVKCRFPVVENGKQISGFGKRFYYKATVMEECDKGFYLDGSDTIVC 112

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     113 DSNSTWDP-PVPKC 125


RESULT 13
US-08-888-171-18
```

```
; Sequence 18, Application US/08888171
; Patent No. 5851528
; GENERAL INFORMATION:
;    APPLICANT: Jane Long, Ko
;    APPLICANT: Higgins, Paul J.
;    APPLICANT: Yeh, C. Grace
;    TITLE OF INVENTION: METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION: ACTIVATION
;    NUMBER OF SEQUENCES: 19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson, P.C.
;      STREET: 225 Franklin Street
;      CITY: Boston
;      STATE: MA
;      COUNTRY: US
;      ZIP: 02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: Windows95
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/888,171
;      FILING DATE: 04-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/310,416
;      FILING DATE: 22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Freeman, John W.
;      REGISTRATION NUMBER: 29,066
;      REFERENCE/DOCKET NUMBER: 06180/005002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 617/542-5070
;      TELEFAX: 617/542-8906
;      TELEX: 200154
;  INFORMATION FOR SEQ ID NO: 18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 169 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-888-171-18

  Query Match              25.1%;  Score 188.5;  DB 2;  Length 169;
  Best Local Similarity    30.6%;  Pred. No. 4.5e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy        1 EISCDAPPEVKNARKPVYSLEVPG-IVLRYTCSPS-   YRLIGEKAIFCISENQVHA 54
            :: | ||| :|| :  :  :  :  : |:| |:   :|  :| | |
Db        1 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV  57

Qy       55 TWDKAPPICESVNKTISCSDPIVPGGPMNKSKAPVHFDCVIFICKANFTMKGNKIVWC 114
             : :| | |    | : |  | |  |    |  : :| | |   | : || |: |
Db       58 WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKAIVMFECDKGFYLDGSDTIVC 112

Qy      115 QANEMWGPTALPVC 128
            :|  | |  :| |
Db      113 DSNSTWDP PVPKC 125


RESULT 14
US-08-310-416A-13
; Sequence 13, Application US/08310416A
; Patent No. 5679546
; GENERAL INFORMATION:
;    APPLICANT: Jane Long Ko et al.
;    TITLE OF INVENTION: "CHIMERIC" PROTEINS WHICH BLOCK
;    TITLE OF INVENTION: COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES: 19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson P.C.
;      STREET: 225 Franklin Street
;      CITY: Boston
;      STATE: Massachusetts
;      COUNTRY: U.S.A.
;      ZIP: 02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;      COMPUTER: IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM: MS-DOS (Version 5.0)
;      SOFTWARE: WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/310,416A
;      FILING DATE: 22-SEP-1994
;      CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Paul I. Clark
;      REGISTRATION NUMBER: 30,162
;      REFERENCE/DOCKET NUMBER: 06180/005001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (617) 542 5070
;      TELEFAX: (617) 542 8906
;      TELEX: 200154
;  INFORMATION FOR SEQ ID NO: 13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 254 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-310-416A-13

  Query Match              25.1%;  Score 188.5;  DB 1;  Length 254;
  Best Local Similarity    30.6%;  Pred. No. 5.9e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy        1 EISCDAPPEVKNARKPVYSLEIVPG-IVLRYTCSPS-   YRLIGEKAIFCISENQVHA 54
            :: | ||| :|| :  :  :  :  : |:| |:   :|  :| | |
Db      125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV  181

Qy       55 TWDKAPPICESVNKTISCSDPIVPGGPMNKSKAPVHFDCVIFICKANFTMKGNKIVWC 114
             : :| | |    | : |  | |  |    |  : :| | |   | : || |: |
Db      182 WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKAIVMFECDKGFYLDGSDTIVC 236

Qy      115 QANEMWGPTALPVC 128
            :|  | |  :| |
Db      237 DSNSTWDP PVPKC 249


RESULT 15
US-08-888-171-13
; Sequence 13, Application US/08888171
; Patent No. 5851528
; GENERAL INFORMATION:
;    APPLICANT: Jane Long, Ko
;    APPLICANT: Higgins, Paul J.
;    APPLICANT: Yeh, C. Grace
;    TITLE OF INVENTION: METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION: ACTIVATION
;    NUMBER OF SEQUENCES: 19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Fish & Richardson, P.C.
;      STREET: 225 Franklin Street
;      CITY: Boston
;      STATE: MA
;      COUNTRY: US
;      ZIP: 02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: Windows95
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/888,171
;      FILING DATE: 04-JUL-1997
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  08/310,416
;       FILING DATE:  22-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Freeman, John W.
;       REGISTRATION NUMBER:  29,066
;       REFERENCE/DOCKET NUMBER:  06180/005002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542 507
;       TELEFAX:  617/542-890
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-888-171-13

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 254;
  Best Local Similarity   30.6%;  Pred. No. 5.9e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | ||| ::|| :  :  : :       : |:| |:      : |||| |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 181

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :  | |    | :  |  |:|  :    |    | :  :| | |    | : || |: |
Db     182  WSRAAPEC-----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     237 DSNSTWDP-PVPKC 249


RESULT 16
US-08-435-149-1
; Sequence 1, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;     APPLICANT:  INNIS, MICHAEL A.
;     APPLICANT:  ZAROR, ISABEL
;     APPLICANT:  CREASEY, ABLA A.
;     TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;     TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;     NUMBER OF SEQUENCES:  26
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CHIRON CORPORATION
;       STREET:  INTELLECTUAL PROPERTY - R440, P.O. BOX 8097
;       CITY:  EMERYVILLE
;       STATE:  CALIFORNIA
;       COUNTRY:  U.S.A.
;       ZIP:  94662-8097
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/435,149
;       FILING DATE:  05-MAY-1995
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SAVEREIDE, PAUL B.
;       REGISTRATION NUMBER:  36,914
;       REFERENCE/DOCKET NUMBER:  0989.001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 601 2585
;       TELEFAX:  (510) 655 3542
;       TELEX:  N/A
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-435-149-1

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 254;
  Best Local Similarity   30.6%;  Pred. No. 5.9e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | ||| ::|| :  :  : :       : |:| |:      : |||| |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 181

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWC 114
             | :  | |    | :  |  |:|  :    |    | :  :| | |    | : || |: |
Db     182 -WSRAAPEC-----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     237 DSNSTWDP-PVPKC 249


RESULT 17
US-08-310-416A-16
; Sequence 16, Application US/08310416A
; Patent No. 5679546
;  GENERAL INFORMATION:
;     APPLICANT:  Jone-Long Ko et al.
;     TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;     TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERAT NG SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Versio- 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/310,415A
;       FILING DATE:  22-SEP-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Paul T. Clark
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  06180/005001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  293 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-310-416A-16

  Query Match             25.1%;  Score  88.5;  DB 1;  Length 293;
  Best Local Similarity   30.6%;  Pred. No. 7.1e-14;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | ||| ::|| :  :  : :       : |:| |:      : |||| |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV- 181
```

```
Qy      55 TWDKAPEICESVNKTISCSGGFVPGGFMNKGSKAPERHGDSVTFICKANEIMKGSKIVWC 114
           | | | |    | : |  |:|  |    |   | :  :| | |   | : || |: |
Db     182  WSKAAPEC    KVVKCRFPVVENGKQISGPGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :  | |  :| |
Db     237 DSNSTWDP PVPKC 249


RESULT 18
US-08-888-171-16
; Sequence 16, Application US/08888171
; Patent No. 5851528
;  GENERAL INFORMATION:
;     APPLICANT:  Jone Louis, Ko
;     APPLICANT:  Higgins, Paul J.
;     APPLICANT:  Yeh, C. Grace
;     TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;     TITLE OF INVENTION:  ACTIVATION
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson, P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/888,171
;       FILING DATE:  03-JUL-1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/310,416
;       FILING DATE:  22-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Freeman, John W.
;       REGISTRATION NUMBER:  29,066
;       REFERENCE/DOCKET NUMBER:  06180/005002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  293 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-888-171-16

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 293;
  Best Local Similarity   30.6%;  Pred. No. 7.1e-13;
  Matches   41;  Conservative   24;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCQPPPEVKNARKFYYSLPIVPG IVLKYTQSPS    - YPIIGEKAIPCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||| |:|  :| |
Db     125 KVLCTPPPKIKNGKHTFSEVEVEFEYLDAVIYSCDPAPGPDDPFSLIGESTIYC-GDNSV  181

Qy      55 TWDKAPEICESVNKTISCSGGFVPGGFMNKGSKAPERHGDSVTFICKANEIMKGSKIVWC 114
           :3  | |    | : |  |:|  |    |   | :  :| | |   | : || |: |
Db     182  WSKAAPEC    KVVKCRFPVVENGKQISGPGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy     115 QANEMWGPTALPVC 128
           :  | |  :| |
Db     237 DSNSTWDP PVPKC 249


RESULT 19
US-08-528-057-46
; Sequence 46, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;     APPLICANT:  PURCELL, Damian F. J.
;     APPLICANT:  RUSSELL, Sarah M.
;     APPLICANT:  McKENZIE, Ian F. C.
;     TITLE OF INVENTION:  CD46 VARIANTS
;     NUMBER OF SEQUENCES:  46
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  3000 K Street, N.W., Suite 500
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20007-5109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  Patentin Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/528,057
;       FILING DATE:  CONCURRENTLY HEREWITH
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/961,686
;       FILING DATE:  11 JAN 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/AU91/00199
;       FILING DATE:  10 MAY 1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PK0133/90
;       FILING DATE:  11 MAY 1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  17222/112 DACO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5400
;       TELEFAX:  (202)672-5399
;       TELEX:  904146
;   INFORMATION FOR SEQ ID NO:  46:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  324 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-528-057-46

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 324;
  Best Local Similarity   30.6%;  Pred. No. 8.1e-13;
  Matches   41;  Conservative   24;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCQPPPEVKNARKFYYSLPIVPG IVLKYTQSPS    - YPIIGEKAIPCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:     : |||| |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVEFEYLDAVIYSCDPAPGPDDPFSLIGESTIYC-GDNSV  215

Qy      55 TWDKAPEICESVNKTISCSGGFVPGGFMNKGSKAPERHGDSVTFICKANEIMKGSKIVWC 114
           :3  | |    | : |  |:|  |    |   | :  :| | |   | : || |: |
Db     216 -WSKAAPEC    KVVKCRFPVVENGKQISGPGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :  | |  :| |
Db     271 DSNSTWDP PVPKC 283


RESULT 20
US-08-528-057-42
; Sequence 42, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;     APPLICANT:  PURCELL, Damian F. J.
;     APPLICANT:  RUSSELL, Sarah M.
;     APPLICANT:  McKENZIE, Ian F. C.
```

```
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,057
;      FILING DATE:  CONCURRENTLY HEREWITH
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,686
;      FILING DATE:  11-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU91/00199
;      FILING DATE:  10-MAY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK0133/90
;      FILING DATE:  11-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  17227/112 DACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  370 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-528-057-42

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 370;
  Best Local Similarity   30.6%;  Pred. No. 9.6e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSIEIVPG-TVLKYTCSPS----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : |||| |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216  WSRAAPEC---KVVKCRFPVVENGKQISGFGKKFYYKAIVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 21
US-08-528-057-44
; Sequence 44, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,057
;      FILING DATE:  CONCURRENTLY HEREWITH
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/961,686
;      FILING DATE:  11-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU91/00199
;      FILING DATE:  10-MAY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PK0133/90
;      FILING DATE:  11-MAY-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  17227/112 DACO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  373 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-528-057-44

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 373;
  Best Local Similarity   30.6%;  Pred. No. 9.7e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSIPIVPG-TV RY TSPS----YRLIGEKAIFCISENQVHA 54
           :: | |:|::|| :  :  : :       : | | |:      : |||| |:|  :| |
Db     159 KVLCPPPKIKNGKHTFSEVEVFEYLDAvTYSCDPAPGPDPFSLIGESTIYC-GDNSV   215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216 -WSRAAPEC----KVVKCRFPVVENGKQISG-GKKFYYKAIVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWEP-PVPKC 283


RESULT 22
US-08-528-057-2
; Sequence 2, Application US/08528057
; Patent No. 5846715
;  GENERAL INFORMATION:
;    APPLICANT:  PURCELL, Damian F. J.
;    APPLICANT:  RUSSELL, Sarah M.
;    APPLICANT:  McKENZIE, Ian F. C.
;    TITLE OF INVENTION:  CD46 VARIANTS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/528,057
;       FILING DATE:  CONCURRENTLY HEREWITH
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/961,686
;       FILING DATE:  11 JAN 1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/AU91/00199
;       FILING DATE:  10 MAY 1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  AU PK0133/90
;       FILING DATE:  11 MAY 1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  17227/112 DAVO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202)672-5300
;       TELEFAX:  (202)672-5399
;       TELEX:  904136
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  377 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-528-057-2

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 377;
  Best Local Similarity   30.6%;  Pred. No. 9.8e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy          1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS -    YRLIGEKAIFCISENQVHA 54
Db        159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYCGDNSV - 215

Qy         55 IWDKAPPICESVNKIISCSDPIVPGGEMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
Db        216 WSRAAPEC    KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy        115 QANEMWGPTALPVC 128
Db        271 DSNSTWDPPVPKC 283


RESULT 23
5514787-2
;Patent No. 5514787
;     APPLICANT: ATKINSON, JOHN P.
;     TITLE OF INVENTION: DNA SEQUENCES ENCODING HUMAN MEMBRANE
;COFACTOR PROTEIN (MCP)
;     NUMBER OF SEQUENCES: 2
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/07/948,350
;       FILING DATE: 21 SEP 1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 384,210
;       FILING DATE: 21 JUL 1989
;SEQ ID NO:2:
;       LENGTH: 384
5514787-2

  Query Match             25.1%;  Score 188.5;  DB 6;  Length 384;
  Best Local Similarity   30.6%;  Pred. No. 1e-12;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy          1 EISCDPPPEVKNARKPYYSLPIVPG TVLRYTCSPS      YRLIGEKAIFCISENQVHA 54
Db        159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYCGDNSV   215

Qy         55 IWDKAPPICESVNKIISCSDPIVPGGEMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114

Db        216 -WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy        115 QANEMWGPTALPVC 128
Db        271 DSNSTWDPPVPKC 283

RESULT 24
US-08-435-149-3
; Sequence 3, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;    APPLICANT:  INNIS, MICHAEL A.
;    APPLICANT:  ZAROR, ISABEL
;    APPLICANT:  CREASEY, ABLA A.
;    TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;    TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  INTELLECTUAL PROPERTY   R440, P.O. BOX 8097
;      CITY:  EMERYVILLE
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,149
;      FILING DATE:  05 MAY 1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAVEREIDE, PAUL B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0989.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2585
;      TELEFAX:  (510) 655-3542
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  577 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-435-149-3

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 577;
  Best Local Similarity   30.6%;  Pred. No. 1.7e-12;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy          1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS -    YRLIGEKAIFCISENQVHA 54
Db        125 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYCGDNSV   181

Qy         55 IWDKAPPICESVNKIISCSDPIVPGGEMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
Db        182 -WSRAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 236

Qy        115 QANEMWGPTALPVC 128
Db        237 DSNSTWDPPVPKC 249


RESULT 25
US-09-475-460A-32
; Sequence 32, Application US/09475460A
; Patent No. 6316253
```

```
; GENERAL INFORMATION:
;  APPLICANT:  Scott, Elizabeth
;  APPLICANT:  Innis, Michael
;  TITLE OF INVENTION:  EXPRESSION VECTORS, TRANSFECTION SYSTEMS, AND METHOD OF
;  TITLE OF INVENTION:  USE THEREOF
;  FILE REFERENCE:  1527.003
;  CURRENT APPLICATION NUMBER: US/09/475,460A
;  CURRENT FILING DATE:  1999-12-30
;  NUMBER OF SEQ ID NOS:  42
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 42
;   LENGTH: 611
;   TYPE: PRT
;   ORGANISM: CAB2
US-09-475-460A-42

  Query Match             25.1%;  Score 188.5;  DB 4;  Length 611;
  Best Local Similarity   30.6%;  Pred. No. 1.8e-12;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 FISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :      : |:| |:      : |||| |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
             | :| | |     | : |  |:|  |     |     | :  :| | |    | : || |: |
Db     216 -WSRAAPEC   --KVVKCPPPVVPNGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 26
US-08-356-361-31
; Sequence 31, Application US/08356361
; Patent No. 5833989
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I.
;    TITLE OF INVENTION:  No. 5833989e1 Compounds
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellecutal Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,361
;      FILING DATE:  03-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H.
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P30423
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5019
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  133 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-356-361-31

  Query Match             23.7%;  Score 178.5;  DB 2;  Length 133;
  Best Local Similarity   30.9%;  Pred. No. 1.1e-12;
  Matches   43;  Conservative   20;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |  || : |        |: ::      |:|: | |:|       : |:|| :|:| |
Db       3 IPCGLPPTITNGDFISTNRENFHY-----GSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 57

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
           :     |    | |   ||   |: | |  | :   ::: |   : | | |: | |||
Db      58 DDQVGIWSGPAPQCTIPNK---CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGP 114

Qy     110 KTVWCQANEMWGPTALPVC 128
           : | |||   | |  || |
Db     115 RRVKCQALNKWEP-ELPSC 132


RESULT 27
US-08-769-967A-31
; Sequence 31, Application US/08769967A
; Patent No. 5859223
;  GENERAL INFORMATION:
;    APPLICANT:  Mossakowska, Danuta E.I.
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman, Anne Mary
;    TITLE OF INVENTION:  Soluble CR1 Derivatives
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,967A
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/440,569
;      FILING DATE:  15-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kinc, William T.
;      REGISTRATION NUMBER:  30,954
;      REFERENCE/DOCKET NUMBER:  P30423C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5364
;      TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  133 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-769-967A-31

  Query Match             23.7%;  Score 178.5;  DB 2;  Length 133;
  Best Local Similarity   30.9%;  Pred. No. 1.1e-12;
  Matches   43;  Conservative   20;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |  || : |        |: ::      |:|: | |:|       : |:|| :|:| |
Db       3 IPCGLPPTITNGDFISTNRENFHY-----GSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 57
```

```
Qy      50 NQVHATWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANPIMKGS 109
           :     |    | |   ||   |: | |  | :   ::: |   : | | | :  | |||
Db      58 DDQVGIWSGPAPQCTIPNK   CTPPNVENGILVSDNRSLPSLNFVVFFRCQPGPVMKGP 114

Qy     110 KTVWCQANPMWGPTALPVC 128
           : | |||   | |  || |
Db     115 RRVKCQALNKWEP-ELPSC 132


RESULT 28
US-08-356-361-29
  Sequence 29, Application US/08356361
; Patent No. 5833989
.  GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
.    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I.
;    TITLE OF INVENTION:  No. 5833989e1 Compounds
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellecutal Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,361
;      FILING DATE:  03-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
.      NAME:  Jervis, Herbert H.
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P3042A
.    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5019
;      TELEFAX:  (610) 270-5090
.  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  254 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-356-361-29

  Query Match             23.7%;  Score 178.5;  DB 2;  Length 254;
  Best Local Similarity   30.9%;  Pred. No. 7.3e-12;
  Matches   44;  Conservative   20;  Mismatches   56;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA       RKPYYSLPIVPGIVLKYICSPS       YRLIGKKAIFCISE 49
           | |  |  : |          : ::       |:|: | |:|       : :|| :|:| |
Db     124 IPCGLPPTLINGDFISTNRENFHY     CSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 178

Qy      50 NQVHATWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANPIMKGS 109
           :     |    | |   ||   |: | |  | :   ::: |   : | | | :  | |||
Db     179 DDQVGIWSGPAPQCTIPNK   CTPPNVENGILVSDNRSLPSLNFVVFFRCQPGPVMKGP 235

Qy     110 KTVWCQANPMWGPTALPVC 128
           : | |||   | |  || |
Db     236 RRVKCQALNKWEP-ELPSC 254


RESULT 29
US-08-769-967A-29
; Sequence 29, Application US/08769967A
; Patent No. 5859223
;  GENERAL INFORMATION:
;    APPLICANT:  Mossakowska, Danuta E.I.
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman, Anne Mary
;    TITLE OF INVENTION:  Soluble CR1 Derivatives
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;      STREET:  P.O. Box 1539
;      CITY:  King of Prussia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,967A
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/440,569
;      FILING DATE:  15-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, William T.
;      REGISTRATION NUMBER:  30,994
;      REFERENCE/DOCKET NUMBER:  P30424C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 270-5364
;      TELEFAX:  (610) 270-5090
.  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  254 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  N-terminal
US-08-769-967A-29

  Query Match             23.7%;  Score 178.5;  DB 2;  Length 254;
  Best Local Similarity   30.9%;  Pred. No. 7.3e-12;
  Matches   44;  Conservative   20;  Mismatches   56;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA       RKPYYSLPIVPGIVLKYICSPS       YRLIGKKAIFCISE 49
           | |  |  : |          : ::       |:|: | |:|       : :|| :|:| |
Db     124 IPCGLPPTLINGDFISTNRENFHY     CSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 178

Qy      50 NQVHATWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANPIMKGS 109
           :     |    | |   ||   |: | |  | :   ::: |   : | | | :  | |||
Db     179 DDQVGIWSGPAPQCTIPNK   CTPPNVENGILVSDNRSLPSLNFVVFFRCQPGPVMKGP 235

Qy     110 KTVWCQANPMWGPTALPVC 128
           : | |||   | |  || |
Db     236 RRVKCQALNKWEP-ELPSC 254


RESULT 30
US-08-356-361-30
; Sequence 30, Application US/08356361
; Patent No. 5833989
;  GENERAL INFORMATION:
;    APPLICANT:  Smith, Richard A.G.
;    APPLICANT:  Dodd, Ian
;    APPLICANT:  Freeman Mary A.
;    APPLICANT:  Mossakowska, Danuta E.I.
;    TITLE OF INVENTION:  No. 5833989e1 Compounds
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporate Intellecutal Property
```

```
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/356,461
;       FILING DATE:  03-Jul-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jervis, Herbert H.
;       REGISTRATION NUMBER:  41,171
;       REFERENCE/DOCKET NUMBER:  P30423
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5019
;       TELEFAX:  (610) 270-5090
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  N-terminal
US-08-356-461-30

  Query Match             23.3%;  Score 175.5;  DB 2;  Length 254;
  Best Local Similarity   30.9%;  Pred. No. 1.5e-11;
  Matches   43;  Conservative   19;  Mismatches   56;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |  || : |         |: ::      |:|: | |:|       : |:|| :|:| |
Db     124 IPCGLPPTITNGDFISTNRENFHY-----GSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 178

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGS 109
           :    |    | |   ||   |: | |  | :   ::: |   : | | |:  | |||
Db     179 DDQVGIWSGPAPQCIIPNK-  CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGP 235

Qy     110 KTVWCQANEMWGPTALPVC 128
             | |||   | |  || |
Db     236 HRVKCQALNKWEP-ELPSC 253


RESULT 31
US-08-769-967A-30
; Sequence 30, Application US/08769967A
; Patent No. 5859223
;  GENERAL INFORMATION:
;     APPLICANT:  Mossakowska, Danuta E.I.
;     APPLICANT:  Smith, Richard A.G.
;     APPLICANT:  Dodd, Ian
;     APPLICANT:  Freeman, Anne Mary
;     TITLE OF INVENTION:  Soluble CRI Derivatives
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/769,967A
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/440,569
;       FILING DATE:  15-May-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  King, William T.
;       REGISTRATION NUMBER:  30,954
;       REFERENCE/DOCKET NUMBER:  P30423C2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5364
;       TELEFAX:  (610) 270-5090
;   INFORMATION FOR SEQ ID NO:  30:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  254 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  N-terminal
US-08-769-967A-30

  Query Match             23.3%;  Score 175.5;  DB 2;  Length 254;
  Best Local Similarity   30.9%;  Pred. No. 1.5e-11;
  Matches   43;  Conservative   19;  Mismatches   56;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |  || : |         |: ::      |:|: | |:|       : |:|| :|:| |
Db     124 IPCGLPPTITNGDFISTNRENFHY-----GSVVTYRCNPGSGGRKVFELVGEPSIYCTSN 178

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGS 109
           :    |    | |   ||   |: | |  | :   ::: |   : | | |:  | |||
Db     179 DDQVGIWSGPAPQCIIPNK-  CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGP 235

Qy     110 KTVWCQANEMWGPTALPVC 128
             | |||   | |  || |
Db     236 HRVKCQALNKWEP-ELPSC 253


RESULT 32
5514582-41
;Patent No. 5514582
;     APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
;     TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;     NUMBER OF SEQUENCES: 43
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/185,670
;       FILING DATE: 21-JAN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 986,931
;       FILING DATE: 08-DEC-1992
;       APPLICATION NUMBER: 808,122
;       FILING DATE: 16-DEC-1991
;       APPLICATION NUMBER: 440,625
;       FILING DATE: 22-NOV-1989
;       APPLICATION NUMBER: 315,015
;       FILING DATE: 23-FEB-1989
;SEQ ID NO:41:
;       LENGTH: 86
5514582-41

  Query Match             21.7%;  Score 163;  DB 6;  Length 86;
  Best Local Similarity   32.6%;  Pred. No. 8.8e-11;
  Matches   29;  Conservative   15;  Mismatches   41;  Indels    4;  Gaps    1;

Qy      32 CSPSYRLIGEKAIFCISENQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFF 91
           | | : |:|  :| | ||:    |  :|| ||     |:|  | |  | |  |  |    :
Db       1 CDPRFSLLGHASISCTVENETIGVWRPSPPTCEK---ITCRKPDVSHGEMVSGFGPIYN 56

Qy      92 HGDSVTFTCKANFTMKGSKTVWCQANEMW 120
           : |:: | |:  | ::||  : | |:  |
Db      57 YKDTIVFKCQKGFVIRGSSVIHCDADSKW 85
```

```
RESULT 63
5514582-98
;Patent No. 5514582
;  APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
;  TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;  NUMBER OF SEQUENCES: 43
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/08/185,670
;    FILING DATE: 21-JAN-1994
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: 986,941
;    FILING DATE: 08-DEC-1992
;    APPLICATION NUMBER: 808,122
;    FILING DATE: 16-DEC-1991
;    APPLICATION NUMBER: 440,625
;    FILING DATE: 22-NOV-1989
;    APPLICATION NUMBER: 415,015
;    FILING DATE: 29-FEB-1989
;SEQ ID NO 98:
;    LENGTH: 124
5514582-98

  Query Match             20.4%;  Score 154.5;  DB 6;  Length 124;
  Best Local Similarity   31.1%;  Pred. No. 1.5e-09;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA---RKPYYSLRIVPGIVLRYDCSPSYRLIGERKAIFQISENQVHATWD 57
           |  |  | : :|     | |||::       : : |  | | |     |    ||:  |
Db       1 IHCPRPHDFENGEYWPRSPYYNV    SDEISFHCYDGYTLRGSANRTC---QVNGRWS 52

Qy      58 KAIPICEEVNKTIKSPDIVDGGFMNKQSKAFFHGDSVIFICKANFTMRGSKTVWCQAN 117
             ||:  |   || | :  |   ||:  :|  |||| |    |::||:   ||
Db      53 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 108

Qy     118 EMWGPTALPVCESDF 132
             |  |  | |:  |
Db     109 GSWSGTE PSCQDSF 122


RESULT 64
US-08-177-109A-57
; Sequence 57, Application US/08177109A
; Patent No. 5869615
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Bourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/177,109A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107 DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-177-109A-57

  Query Match             20.4%;  Score 154.5;  DB 2;  Length 265;
  Best Local Similarity   31.1%;  Pred. No. 4.1e-09;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA---RKPYYSLRIVPGIVLRYDCSPSYRLIGERKAIFQISENQVHATWD 57
           |  |  | : :|     | |||::       : : |  | | |     |    ||:  |
Db      76 IHCPRPHDFENGEYWPRSPYYNV    SDEISFHCYDGYTLRGSANRTC    QVNGRWS 127

Qy      58 KAIPICEEVNKTIKSPDIVDGGFMNKQSKAFFHGDSVIFICKANFTMRGSKTVWCQAN 117
             ||:  |   || | :  |   ||:  :|  |||| |    |::||:   ||
Db     128 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 183

Qy     118 EMWGPTALPVCESDF 132
             |  |  | |:  |
Db     184 GSWSGTE PSCQDSF 197


RESULT 65
US-08-687-706-57
; Sequence 57, Application US/08687706
; Patent No. 5928892
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Bourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,706
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/177,109
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107 DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-687-706-57

  Query Match             20.4%;  Score 154.5;  DB 2;  Length 265;
  Best Local Similarity   31.1%;  Pred. No. 4.1e-09;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;
```

```
Qy       2 ISCDPPPEVKNA    RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           | |  | : :|     | |||::       : : |   | | |     |    ||:  |
Db      76 IHCPRPHDFENGEYWPRSPYYNV----SDEISFHCYDGYTLRGSANRTC----QVNGRWS 127

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             ||:  |    ||:| :| |    ||:  :|  |||| |    |::||:   ||
Db     128 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 183

Qy     118 EMWGPTALPVCESDF 132
           | |  | |:  |
Db     184 GSWSGTE-PSCQDSF 197


RESULT 36
US-08-177-109A-2
; Sequence 2, Application US/08177109A
; Patent No. 5869615
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/177,109A
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-177-109A-2

  Query Match             20.4%;  Score 153.5;  DB 2;  Length 764;
  Best Local Similarity   31.1%;  Pred. No. 1.6e-08;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA----RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           | |  | : :|     | |||::       : : |   | | |     |    ||:  |
Db     101 IHCPRPHDFENGEYWPRSPYYNV---SDEISFHCYDGYTLRGSANRTC----QVNGRWS 152

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             ||:  |    ||:| :| |    ||:  :|  |||| |    |::||:   ||
Db     153 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 208

Qy     118 EMWGPTALPVCESDF 132
           | |  | |:  |
Db     209 GSWSGTE-PSCQDSF 222


RESULT 37
US-08-687-706-2
; Sequence 2, Application US/08687706
; Patent No. 5928892
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;    TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  USA
;      ZIP:  30309-3450
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,706
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/177,109
;      FILING DATE:  03-JAN-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pabst, Patrea L.
;      REGISTRATION NUMBER:  31,284
;      REFERENCE/DOCKET NUMBER:  WU 107 DIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (404) 873-8794
;      TELEFAX:  (404) 873-8795
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  764 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
US-08-687-706-2

  Query Match             20.4%;  Score 153.5;  DB 2;  Length 764;
  Best Local Similarity   31.1%;  Pred. No. 1.6e-08;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA----RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           | |  | : :|     | |||::       : : |   | | |     |    ||:  |
Db     101 IHCPRPHDFENGEYWPRSPYYNV----SDEISFHCYDGYTLRGSANRTC----QVNGRWS 152

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             ||:  |    ||:| :| |    ||:  :|  |||| |    |::||:   ||
Db     153 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 208

Qy     118 EMWGPTALPVCESDF 132
           | |  | |:  |
Db     209 GSWSGTE-PSCQDSF 222


RESULT 38
5514582-35
;Patent No. 5514582
;    APPLICANT: CAPON, DANIEL J.;LASKY, LAURENCE A.
;    TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;    NUMBER OF SEQUENCES: 43
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/185,670
;      FILING DATE: 21-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 986,931
```

```
;       FILING DATE: 08-DEC-1992
;       APPLICATION NUMBER: 808,122
;       FILING DATE: 16-DEC-1991
;       APPLICATION NUMBER: 440,625
;       FILING DATE: 22-NOV-1989
;       APPLICATION NUMBER: 415,015
;       FILING DATE: 23-FEB-1989
;SEQ ID NO: 43;
;       LENGTH: 126
5514582.43

  Query Match              20.0%;  Score 150.5;  DB 6;  Length 126;
  Best Local Similarity    60.9%;  Pred. No. 3.3e-09;
  Matches   44;  Conservative   16;  Mismatches   55;  Indels   25;  Gaps    8;

Qy        4 CDPPPEVKNAR      KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDK 58
            |||:|  |         |  :|    ||  : |:| | | | |  :: |  :       |
Db        1 CQPPPDVLHAERTQRDKDNTS    PGQEVFYSCEPGYDLRGAASMRCTPQGD  - WSP 52

Qy       59 APPICESVNKTISCSDPIVPGGFMNKGSKAP  FRHGDSVTFTCKANFTMKGSKTVWC-  114
            | : ||       ||  |    |  :|     |  : |  | | |    | :|||   :|
Db       53 AAPICE    VKSCDD  FMGQLLNGRVLFFVNLQLGAKVDFVCDEGFQLKGSSASYCVL 106

Qy      115  QANEMWGPTALPVCESDE 132
               :|  ::::|||  |
Db      107 AGMESLWN SSVPVCEQIF 124


RESULT 43
5514582.43
;Patent No. 5514582;
;     APPLICANT: CARON, DANIEL J.;LASKY, LAURENCE A.
;     TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
;IMMUNOGLOBULINS
;     NUMBER OF SEQUENCES: 13
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/185,670
;       FILING DATE: 21-JAN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 986,941
;       FILING DATE: 08-DEC-1992
;       APPLICATION NUMBER: 808,122
;       FILING DATE: 16-DEC-1991
;       APPLICATION NUMBER: 440,625
;       FILING DATE: 22-NOV-1989
;       APPLICATION NUMBER: 415,015
;       FILING DATE: 23-FEB-1989
;SEQ ID NO:43;
;       LENGTH: 126
5514582.43

  Query Match              19.8%;  Score 145.5;  DB 6;  Length 126;
  Best Local Similarity    26.2%;  Pred. No. 1.2e-08;
  Matches   44;  Conservative   22;  Mismatches   62;  Indels    9;  Gaps    3;

Qy        4 CDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPPIC 63
            |  |  :     :     :  : |:|:  | ||||   :|   :       |:   |||
Db        5 CPSPRDITNGQLIIYGGVIF  GSSTIYSCNSGYHLIGESKSVCFICGSTGSMVWNPEAPIC 62

Qy       64 ESVNKTISCSDFIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWGPT 123
            |||     |  |       :  :  :  | ||:|| : ::: |:  |  |     |
Db       63 ESV     KCQSFPSISNGRHNGYFDFYIIGSVVIYSCNSGYSLIGNSGVICSGGFWSDP 117

Qy      124 ALPVCE 129
               ||
Db      118   PTCQ 121


RESULT 44
US-08-456-361-27
; Sequence 27, Application US/08456361
; Patent No. 5843989
;  GENERAL INFORMATION:
;     APPLICANT:  Smith, Richard A.G.
;     APPLICANT:  Dodd, Ian
;     APPLICANT:  Freeman Mary A.
;     APPLICANT:  Mossakowska, Danuta E.I.
;     TITLE OF INVENTION:  No. 5843989CI Compounds
;     NUMBER OF SEQUENCES:  44
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/456,361
;       FILING DATE:  03-Jul-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Jervis, Herbert H.
;       REGISTRATION NUMBER:  31,171
;       REFERENCE/DOCKET NUMBER:  P30423
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5019
;       TELEFAX:  (610) 270-5090
;  INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  197 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  N-terminal
US-08-456-361-27

  Query Match              19.1%;  Score 144;  DB 2;  Length 197;
  Best Local Similarity    29.5%;  Pred. No. 3.1e-08;
  Matches   41;  Conservative   22;  Mismatches   54;  Indels   22;  Gaps    9;

Qy        3 SCDPPPEVKNAPKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPPI 62
            ||  ||:  |    :    |  |: :::|:|:  |||||  :  ||       |   ||
Db       63 SCRNPPDPVNG-MVHVIKGIQFGSQIKYSCIKGYRLIGSSSAICLISGDTEVIWDNEIPI 120

Qy       63 CESVNKTISCSDF---IVPGGFMNKGSKAPFRHGDSVTFTCKAN----- FTMKGSKTVWC 114
            |:     | |  |  |  | |::  ::  | :|  ||: |           | : |  :::|
Db      121 CD---  RIPCGLPPTITNGDFIST NRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYC 175

Qy      115 QANE---  MW GPIALPVC 128
              :|:    :: ||   | |
Db      176 TSNDDQVGIWSGPA--PQC 192


RESULT 41
US-08-769-967A-27
; Sequence 27, Application US/08769967A
; Patent No. 5859228
;  GENERAL INFORMATION:
;     APPLICANT:  Mossakowska, Danuta E.I.
;     APPLICANT:  Smith, Richard A.G.
;     APPLICANT:  Dodd, Ian
;     APPLICANT:  Freeman, Anne Mary
;     TITLE OF INVENTION:  Soluble CR1 Derivatives
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  SmithKline Beecham Corporate Intellectual Property
;       STREET:  P.O. Box 1539
;       CITY:  King of Prussia
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19406
```

```
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC DOS/MS DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/769,967A
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/440,569
;       FILING DATE:  15 May 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  King, William T.
;       REGISTRATION NUMBER:  30,954
;       REFERENCE/DOCKET NUMBER:  P3042C2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (610) 270-5464
;       TELEFAX:  (610) 270-5090
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  197 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  N-terminal
US-08-769-967A-27

  Query Match             19.1%;  Score 144;  DB 2;  Length 197;
  Best Local Similarity   29.5%;  Pred. No. 3.1e-08;
  Matches   41;  Conservative   22;  Mismatches   54;  Indels   22;  Gaps    9;

Qy       8 SCDPPPFVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPPI 62
           ||  ||:  |    :    |  |: ::|:|:  ||||  :  ||          ||   ||
Db      63 SCRNPPDPVNG MVHVIKGIQPGSQIKYSCTKGYRLIGSSSATCIISGDT-VIWDNETPI 120

Qy      63 CESVNKTISCSDP--IVIGGFMNKGSKAPFRHGDSVTFTCKAN------FTMKGSKTVWC 114
           |:      | |  |  |  | |::  ::  | :|  ||: |           | : |  :::|
Db     121 CD--RIPCGLPPTITNGDFIST-NRENFHYGSVVTYRCNPGSGGRKVFELVGEPSIYC 175

Qy     115 QANE---MW GPTALPVC 128
             :|:    :| ||   | |
Db     176 ISNDDQVGIWSGPA  PQC 192


RESULT 42
US-08-435-149-2
; Sequence 2, Application US/08435149
; Patent No. 5866402
;  GENERAL INFORMATION:
;    APPLICANT:  INNIS, MICHAEL A.
;    APPLICANT:  ZAROR, ISABEL
;    APPLICANT:  CREASEY, ARLA A.
;    TITLE OF INVENTION:  CHIMERIC MCP AND DAF PROTEINS WITH CELL
;    TITLE OF INVENTION:  SURFACE LOCALIZING DOMAIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHIRON CORPORATION
;      STREET:  INTELLECTUAL PROPERTY - R440, P.O. BOX 8097
;      CITY:  EMERYVILLE
;      STATE:  CALIFORNIA
;      COUNTRY:  U.S.A.
;      ZIP:  94662-8097
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC DOS/MS DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/435,149
;      FILING DATE:  05-MAY-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAVEREIDE, PAUL B.
;      REGISTRATION NUMBER:  36,914
;      REFERENCE/DOCKET NUMBER:  0989.C01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2585
;      TELEFAX:  (510) 655-3542
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  323 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-435-149-2

  Query Match             18.9%;  Score 142;  DB 2;  Length 323;
  Best Local Similarity   29.9%;  Pred. No. 9.6e-08;
  Matches   40;  Conservative   24;  Mismatches   50;  Indels   20;  Gaps    9;

Qy       3 SCDPPPFVKNARKPYYSLP--IVPGTVLRYTCSPSYRLIGEKAIFC ISENQVHATWDKA 59
           ||  | |:::| :    :|  |: |  : ::::  |:| |  : || || : |   |
Db     128 SCPNPGEIRNGQ---IDVPGGILFGATISFSVNIGYKLFGSTSSFCLISGSSVQ- WSDP 182

Qy      60 PPICESVNKTISC-SDPIVPGGFMNKGSKAP-RHGDSVTFTCKANFIMKGSKIVWCQAN 117
           | |    : | | : |  | : :|    : :  |::: |   ||| |  :::|  |
Db     183 LPEC----REIYCPAPPQIDNGII-QGERDHYGYRQSVTYACNKGFTMIGEHSIYCIVNN 237

Qy     118 ---EMWGPTALPVC 128
               |   |   | |
Db     238 DEGEWSGPP  PEC 249


RESULT 43
US-08-310-416A-14
; Sequence 14, Application US/08310416A
; Patent No. 5679546
;  GENERAL INFORMATION:
;    APPLICANT:  Jone-Long Ko et al.
;    TITLE OF INVENTION:  CHIMERIC PROTEINS WHICH BLOCK
;    TITLE OF INVENTION:  COMPLEMENT ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,416A
;      FILING DATE:  22-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul T. Clark
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  0618.0/CC5001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  324 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
US-08-810-416A-14

  Query Match             18.9%;  Score 142;  DB 1;  Length 424;
  Best Local Similarity   29.9%;  Pred. No. 9.6e-08;
  Matches   40;  Conservative   24;  Mismatches   50;  Indels   20;  Gaps    9;

Qy        3 SCQPPPEVKNARKPYYSLP  IVNGIVLKYICSPSSYRLIGEKAIFC ISENQVHAIWDKA 59
            ||  | |:::| :      :|  | |  | : :::|  |:| |  : || || : |   |
Db      128 SCPNPGEIRNGQ   IDVPGGILEGAIISFSCNIGYKLFGSISSFCLISGSSVQ  WSDP 182

Qy       60 PPICESVNKIISC SGQIVPGGFMNKGSKAFFHGDSVIFICKANFIMKGSKIVWCQAN- 117
             | |    : | | :  |  : :| :   :  :  |||: |   ||| |  :::|  |
Db      183 LPEC    REIYCPAPPQIDNGII QGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVNN 237

Qy      118 -  EMWGPTALPVC 128
               |  ||   | |
Db      238 DEGEWSGPP  PEC 249


RESULT 44
US-08-888-171-14
; Sequence 14, Application US/08888171
; Patent No. 5851528
;  GENERAL INFORMATION:
;    APPLICANT:  Jone Long, Ko
;    APPLICANT:  Higgins, Paul J.
;    APPLICANT:  Yeh, C. Grace
;    TITLE OF INVENTION:  METHODS OF INHIBITING COMPLEMENT
;    TITLE OF INVENTION:  ACTIVATION
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,171
;      FILING DATE:  03-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/810,416
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06180/005002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  424 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-888-171-14

  Query Match             18.9%;  Score 142;  DB 2;  Length 424;
  Best Local Similarity   29.9%;  Pred. No. 9.6e-08;
  Matches   40;  Conservative   24;  Mismatches   50;  Indels   20;  Gaps    9;

Qy        3 SCQPPPEVKNARKPYYSLP  IVNGIVLKYICSPSSYRLIGEKAIFC ISENQVHAIWDKA 59
            ||  | |:::| :      :|  | |  | : :::|  |:| |  : || || : |   |
Db      128 SCPNPGEIRNGQ   IDVPGGILEGAIISFSCNIGYKLFGSISSFCLISGSSVQ  WSDP 182

Qy       60 PPICESVNKIISC SGQIVPGGFMNKGSKAFFHGDSVIFICKANFIMKGSKIVWCQAN- 117
             | |    : | | :  |  : :| :   :  :  |||: |   ||| |  :::|  |
Db      183 LPEC    REIYCPAPPQIDNGII QGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVNN 237

Qy      118 -  EMWGPTALPVC 128
               |  ||   | |
Db      238 DEGEWSGPP  PEC 249


RESULT 45
US-08-365-470-3
; Sequence 3, Application US/08365470
; Patent No. 5632991
;  GENERAL INFORMATION:
;    APPLICANT:  Gimbrone, Jr., Michael A.
;    TITLE OF INVENTION:  Antibodies Specific For E-Selectin And The Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX
;      STREET:  1100 New York Ave., NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/365,470
;      FILING DATE:  herewith
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/132,510
;      FILING DATE:  05-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/850,802
;      FILING DATE:  13-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Markowicz, Karen R.
;      REGISTRATION NUMBER:  36,351
;      REFERENCE/DOCKET NUMBER:  0627.1450003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  610 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-365-470-3

  Query Match             18.5%;  Score 139;  DB 1;  Length 610;
  Best Local Similarity   31.5%;  Pred. No. 4.6e-07;
  Matches   34;  Conservative   18;  Mismatches   36;  Indels   20;  Gaps    6;

Qy       30 YICSPSYRLIGEKAIFCISENQVHAIWDKAPPICESVNKIISCSQIVPGGFMN  85GSK 87
              :||   : | |   : | :: |   | :  |:||:   | : |:|    |:||   :
Db      433 PTCEEGFMLQGPAQVECTTQGQ-- WTQQIPVCEAFQCT ALSNP- ERGYMNCLPSAS 485

Qy       88 APFPHCDSVITPTCKANFTMFGSKTVWCQANEMWGPTA     LPVCES 130
            |:|:| | |:|:  | :|||| : |      |||       : ||:
Db      486 GSERYGSSCEFSCPQGFVLKGSKRTQC      GPTGEWDNEKPTCEA 527


Search completed: November 6, 2002, 16:06:36
Job time : 13.549 secs
```

GenCore version 5.1.3

OM protein protein search, using sw model

Run on: November 6, 2002, 16:04:59 ; Search time 14.5539 Seconds
(without alignments)
891.311 Million cell updates/sec

Title: US-09-834-309-6
Perfect score: 752
Sequence: 1 EISCDPPPEVKNARKPYYSL.....ANEMWGPTALPVCESDFPLE 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	752	100.0	1025	1	A44526	complement C3d/Eps
2	749	99.6	463	2	B45900	complement C3d/Eps
3	749	99.6	676	2	A45900	complement C3b rec
4	466.5	62.0	1091	1	PL0009	complement C3d/Eps
5	232	30.9	597	1	S54711	C4BP alpha chain p
6	225.5	30.0	597	1	NBHUC4	C4b-binding protei
7	222	29.5	469	1	NBMSC4	C4b-binding protei
8	221.5	29.5	2014	2	I36936	complement recepto
9	221.5	29.5	2489	2	I73012	complement C3b/C4b
10	215	28.6	579	2	A56740	sperm-egg recognit
11	208.5	27.7	558	2	S57953	C4BP protein alpha
12	200	26.6	610	1	I46001	C4b-binding protei
13	197.5	26.3	263	1	C36838	complement control
14	197.5	26.3	263	2	T28450	hypothetical prote
15	195.5	26.0	263	2	B72152	B18L protein - var
16	194.5	25.7	263	1	WMVZSP	apolipoprotein H h
17	188.5	25.1	449	2	G02913	sperm CD46 - human
18	188.5	25.1	469	2	I57998	membrane cofactor
19	188.5	25.1	477	2	I54479	membrane cofactor
20	188.5	25.1	484	2	S01896	membrane cofactor
21	183.5	24.1	462	2	JC5194	membrane cofactor
22	181.5	24.1	469	2	JC5148	membrane cofactor
23	176	23.4	1244	1	NBMSH	complement factor
24	172.5	22.9	497	2	JC2054	complement regulat
25	168	22.3	808	2	D45969	complement factor
26	167	22.2	440	2	A43519	complement recepto
27	164.5	21.9	482	2	A34924	complement C3b/C4b
28	157.5	20.9	669	2	S65551	factor H - bovine
29	156.5	20.8	252	2	A34877	C4b-binding protei
30	153.5	20.4	764	1	BBHU	complement factor
31	150	19.9	330	2	I55975	X/Y protein - mous
32	149	19.8	1231	1	NBHUH	complement factor
33	148.5	19.7	551	2	I46709	endothelial leukoc
34	144.5	19.2	152	2	A53274	complement factor
35	143.5	19.1	768	2	A42755	P-selectin precurs
36	142.5	18.9	302	1	WMBE1E	secretory compleme
37	142.5	18.9	360	1	WMBE2E	membrane bound com
38	142	18.9	381	1	B26359	decay-accelerating
39	142	18.9	440	2	A26359	decay-accelerating
40	142	18.9	485	2	S36772	E-selectin - bovin
41	139.5	18.6	343	2	G35070	apolipoprotein H-r
42	139.5	18.6	360	2	T42921	complement control
43	139	18.5	610	2	A35046	E-selectin precurs
44	138.5	18.4	977	2	I52657	seizure-related pr
45	137	18.2	661	1	KFHU13	coagulation factor
46	135	18.0	668	2	A46013	coagulation factor
47	134	17.8	297	1	NBRT	apolipoprotein H p
48	133.5	17.8	258	2	S57960	C4BP protein beta
49	131.5	17.5	612	2	B42755	E-selectin precurs
50	131.5	17.5	2043	2	T18524	scavenger receptor
51	130.5	17.4	752	1	C2HU	complement C2 prec
52	130.5	17.4	761	1	BBMS	complement factor
53	130.5	17.4	1019	2	A38738	coagulation factor
54	130	17.3	340	2	I56234	decay-accelerating
55	129.5	17.2	768	2	I53821	P-selectin - rat
56	127	16.9	345	1	NBBO	apolipoprotein H p
57	126	16.8	764	2	I50807	complement factor
58	125.5	16.7	330	2	I56100	complement factor
59	124.5	16.6	646	2	JN0473	P-selectin precurs
60	124	16.5	345	1	NBMS	apolipoprotein H p
61	123.5	16.4	482	2	JC5092	E-selectin - pig
62	123	16.4	345	1	JN0465	apolipoprotein H p
63	122.5	16.3	1797	2	T21889	hypothetical prote
64	122.5	16.3	1805	2	T21888	hypothetical prote
65	121	16.1	345	1	NBHU	apolipoprotein H p
66	120.5	16.0	560	2	T16833	hypothetical prote
67	119	15.8	830	2	A30359	P-selectin precurs
68	118	15.7	1506	2	T30886	integumentary muci
69	117.5	15.6	202	1	A44247	C4b-binding protei
70	117	15.6	760	1	C2MS	classical compleme
71	116.5	15.5	747	2	I51579	complement factor
72	116	15.4	270	2	I37278	complement factor
73	113.5	15.1	843	1	A27340	complement C7 prec
74	112.5	15.0	303	2	H35068	apolipoprotein H r
75	108.5	14.4	317	2	JQ1799	B5R protein precur
76	107	14.2	1827	2	T34288	hypothetical prote
77	106.5	14.2	1053	2	S46199	probable complemen
78	104.5	13.9	317	2	A38493	ps/hr protein - va
79	104.5	13.9	449	1	NBHUHS	complement factor
80	104.5	13.9	934	1	A34372	complement C6 prec
81	102.5	13.6	317	2	D42526	B5R protein - vacc
82	101.5	13.5	702	2	T16832	hypothetical prote
83	101.5	13.5	2824	2	T22759	hypothetical prote
84	100.5	13.4	317	2	F72172	B7R protein - vari
85	100.5	13.4	317	2	G36855	B7R protein - vari
86	99.5	13.2	317	2	T28605	hypothetical prote
87	99.5	13.2	331	2	A45222	complement factor
88	98.5	13.1	317	2	T37442	EEV gp42.ps/hr pro
89	95.5	12.7	914	1	JN0550	iodide peroxidase
90	95.5	12.7	914	1	S07047	iodide peroxidase
91	94.5	12.6	372	2	S23936	L-selectin precurs
92	94.5	12.6	452	2	A35068	complement factor
93	93.5	12.4	933	1	OPHUIT	iodide peroxidase
94	88.5	11.8	372	2	JC5377	L-selectin precurs
95	87.5	11.6	868	2	T20239	hypothetical prote
96	87.5	11.6	926	1	OPPGIT	iodide peroxidase
97	85	11.3	198	2	I46002	C4BP beta chain
98	84.5	11.2	370	2	S22124	L-selectin precurs
99	84.5	11.2	385	1	A34015	L-selectin precurs
100	83	11.0	347	1	HPMS	haptoglobin precur

ALIGNMENTS

RESULT 1
A45526
complement C3d/Epstein Barr virus receptor 2 precursor - mouse
N;Alternate names: complement receptor type 2
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A45526; A45538; A32215; A45802; B32215
R;Finneroth, J.D.
J. Immunol. 144, 4458-4467, 1990
A;Title: Comparative structure and evolution of murine CR2, the homolog of the human C3d
A;Reference number: A45526; MUID:90229735
A;Accession: A45526
A;Molecule type: mRNA
A;Residues: 1-1025 <FIN>
A;Cross-references: GB:M35684; EMBL:J04153; NID:g192687; PIDN:AAA37448.1; PID:g192688
R;Molina, H.; Kinoshita, T.; Inoue, K.; Carel, J.C.; Holers, V.M.
J. Immunol. 145, 2974-2983, 1990
A;Title: A molecular and immunochemical characterization of mouse CR2. Evidence for a si
A;Reference number: A45538; MUID:91010789
A;Accession: A45538
A;Molecule type: mRNA
A;Residues: 12-305,'I',307-519, A',521-1025 <MOL>
A;Cross-references: GB:M61142; NID:g192692; PIDN:AAA63295.1; PID:g192693
R;Finnerofh, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989
A;Title: Identification of murine complement receptor type 2.
A;Reference number: A32215; MUID:89098890
A;Accession: A32215
A;Molecule type: mRNA
A;Residues: 443-481;991-1025 <F.2>
A;Cross-references: GB:J04153
R;Kurtz, C.B.; Paul, M.S.; Aeperter, M.; Weis, J.J.; Weis, J.H.
J. Immunol. 143, 2058-2067, 1989
A;Title: Murine complement receptor gene family. Identification and characterization of
A;Reference number: A45802; MUID:89381450
A;Accession: A45802
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-'I',100-181,292-961,964-1025 <KUR>
A;Cross-references: GB:M29281; NID:g192685; PIDN:AAA37447.1; PID:g387131
A;Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for residue
A;Note: the authors translated the codon CAC for residue 727 as Asn
C;Superfamily: complement C3d/Epstein Barr virus receptor; complement factor I repeat ho
C;Keywords: alternative splicing; duplication; glycoprotein; receptor; transmembrane pro
F;1-11/Domain: signal sequence #status predicted <SIG>
F;12-1025/Product: complement C3d/Epstein Barr virus receptor 2 (15-repeat form) #status
F;12-973/Domain: extracellular #status predicted <EXT>
F;14-73/Domain: complement factor H repeat homology <FH01>
F;82-138/Domain: complement factor H repeat homology <FH02>
F;146-202/Domain: complement factor H repeat homology <FH03>
F;207-263/Domain: complement factor H repeat homology <FH04>
F;268-324/Domain: complement factor H repeat homology <FH05>
F;333-389/Domain: complement factor H repeat homology <FH06>
F;402-458/Domain: complement factor H repeat homology <FH07>
F;463-514/Domain: complement factor H repeat homology <FH08>
F;519-568/Domain: complement factor H repeat homology <FH09>
F;584-643/Domain: complement factor H repeat homology <FH10>
F;654-704/Domain: complement factor H repeat homology <FH12>
F;709-759/Domain: complement factor H repeat homology <FH13>
F;778-834/Domain: complement factor H repeat homology <FH14>
F;841-897/Domain: complement factor H repeat homology <FH15>
F;902-958/Domain: complement factor H repeat homology <FH16>
F;968-989/Domain: transmembrane #status predicted <TMM>
F;990-1025/Domain: intracellular #status predicted <INT>

Query Match 100.0%; Score 752; DB 1; Length 1025;
Best Local Similarity 100.0%; Pred. No. 6.9e-64;
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EISCDPPPEVKNAKKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
Db 11 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 70

Qy 61 PICESVNKTISCPSPIVPVGEMNKCSKAPPKECISVTFTCKANFTMKGCSKTVWCLANEMW 120
Db 71 PICESVNKTISCPSPIVPVGEMNKCSKAPPKECISVTFTCKANFTMKGCSKTVWCLANEMW 130

Qy 121 GPTALPVCESDFPLE 135
Db 131 GPTALPVCESDFPLE 145

RESULT 2
B45900
complement C3d/Epstein Barr virus receptor precursor alternative splice form - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45900; C45900
R;Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 4581-4591, 1990
A;Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A;Reference number: A45900; MUID:90229754
A;Accession: B45900
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-36 <KUR>
A;Accession: C45900
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-19,'E',84-363 <KU2>
C;Superfamily: complement C3d/Epstein Barr virus receptor; complement factor H repeat
F;24-81/Domain: complement factor H repeat homology <FH1>
F;86-145/Domain: complement factor H repeat homology <FH2>
F;154-210/Domain: complement factor H repeat homology <FH3>
F;218-274/Domain: complement factor H repeat homology <FH4>
F;279-335/Domain: complement factor H repeat homology <FH5>

Query Match 99.6%; Score 749; DB 2; Length 363;
Best Local Similarity 99.3%; Pred. No. 4.5e-64;
Matches 144; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EISCDPPPEVKNAKKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
Db 83 QISCDPPPEVKNAKKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 142

Qy 61 PICESVNKTISCPSPIVPVGEMNKCSKAPPKECISVTFTCKANFTMKGCSKTVWCLANEMW 120
Db 143 PICESVNKTISCPSPIVPVGEMNKCSKAPPKECISVTFTCKANFTMKGCSKTVWCLANEMW 202

Qy 121 GPTALPVCESDFPLE 135
Db 203 GPTALPVCESDFPLE 217

RESULT 3
A45900
complement C4b receptor type 2 long form precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Oct-1999
C;Accession: A45900; 148306
R;Kurtz, C.B., O'Toole, E., Christensen, S.M.; Weis, J.H.
J. Immunol. 144, 3581-3591, 1990
A;Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2
A;Reference number: A45900; MUID:90229754
A;Accession: A45900
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-676 <KUR>
A;Cross-references: GB:M36470
A;Experimental source: clone 31-1
R;Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers,
J. Exp. Med. 181, 151-159, 1995
A;Title: Mouse complement regulatory protein Crry/p65 uses the specific mechanisms of

```
A;Reference number: I48406; MUID:95105691
A;Accession: I48406
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 21-667 <RES>
A;Cross-references: EMBL:U17128; NID:g595980; PIDN:AAA78271.1; PID:g595982
C;Genetics:
A;Gene: CR2
A;Introns: 80/2  113/3  142/2  279/2  633/1
C;Superfamily: complement factor H repeat homology
F;22-78/Domain: complement factor H repeat homology <FH01>
F;83-140/Domain: complement factor H repeat homology <FH02>
F;145-211/Domain: complement factor H repeat homology <FH03>
F;217-272/Domain: complement factor H repeat homology <FH04>
F;276-341/Domain: complement factor H repeat homology <FH05>
F;346-394/Domain: complement factor H repeat homology <FH06>
F;399-458/Domain: complement factor H repeat homology <FH07>
F;467-523/Domain: complement factor H repeat homology <FH08>
F;531-587/Domain: complement factor H repeat homology <FH09>
F;592-648/Domain: complement factor H repeat homology <FH10>

  Query Match             99.6%;  Score 749;  DB 2;  Length 676;
  Best Local Similarity   99.3%;  Pred. No. 8.7e-64;
  Matches  134;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 FISCDPPPFVKNARKPYYSLPIVPGTVLRYTCSPSYPLIGEKAIFCISENQVHATWDKAP 60
           :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 QISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 455

Qy      61 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQANEMW 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     456 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 515

Qy     121 GPTALPVCESDFPLE 135
           |||||||||||||||
Db     516 GPTALPVCESDFPLE 530


RESULT 4
PL0009
Complement C3d/Epstein-Barr virus receptor precursor - human
N;Alternate names: complement receptor 2; CR2/CD21
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1992 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C;Accession: JL0028; A39958; A32036; A24319; B24319; C24319; D24319; E24319; PL0
R;Weis, J.J.; Toothaker, L.E.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
J. Exp. Med. 167, 1047-1066, 1988
A;Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus
A;Reference number: JL0028; MUID:88171282
A;Accession: JL0028
A;Molecule type: mRNA
A;Residues: 1-1091 <WEI>
A;Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542 hav
R;Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987
A;Title: Molecular cloning of the cDNA encoding the Epstein-Barr virus/C3d receptor (com
A;Reference number: A39958; MUID:88097454
A;Accession: A39958
A;Molecule type: mRNA
A;Residues: 1-456,'G',457-644,'R',646-669,'R',671-816,'NCSAEVILKAWILERAF',835-840,'L',84
A;Cross-references: GB:M29565; NID:g181919; PIDN:AAA35784.1; PID:g181920
R;Fujisaku, A.; Harley, J.B.; Frank, M.B.; Gruner, B.A.; Frazier, B.; Holers, V.M.
J. Biol. Chem. 264, 2118-2125, 1989
A;Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus rece
A;Reference number: A32036; MUID:89123277
A;Accession: A32036
A;Molecule type: mRNA
A;Residues: 1-456,'G',457-658,718-1050,'I',1052-1060,'E',1062-1091 <FUJ>
A;Cross-references: GB:J04463
R;Weis, J.J.; Fearon, D.T.; Klickstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Kops,
Proc. Natl. Acad. Sci. U.S.A. 83, 5639-5643, 1986
A;Title: Identification of a partial cDNA clone for the C3d/Epstein-Barr virus receptor
 of complement.

A;Reference number: A94114; MUID:86287311
A;Accession: A24319
A;Molecule type: protein
A;Residues: 226-230,'XILQ';257-267;332-334;584-591,'Q',593,'S',595-596;728-735-WEI>
A;Experimental source: B-lymphoblastoid cell lines SB and Raji
C;Genetics:
A;Gene: GDB:CR2
A;Cross-references: GDB:119802; OMIM:120650
A;Map position: 1q32-q42
C;Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat
C;Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-1091/Product: complement receptor 2 (16 repeat form) #status predicted <MATU>
F;21-658 718-1091/Product: complement receptor 2 (15 repeat form) #status predicted <
F;23-82/Domain: complement factor H repeat homology <FH01>
F;91-146/Domain: complement factor H repeat homology <FH02>
F;154-210/Domain: complement factor H repeat homology <FH03>
F;215-271/Domain: complement factor H repeat homology <FH04>
F;276-342/Domain: complement factor H repeat homology <FH05>
F;351-406/Domain: complement factor H repeat homology <FH06>
F;410-465/Domain: complement factor H repeat homology <FH07>
F;470-521/Domain: complement factor H repeat homology <FH08>
F;526-592/Domain: complement factor H repeat homology <FH09>
F;601-656/Domain: complement factor H repeat homology <FH10>
F;660-716/Domain: complement factor H repeat homology <FH11>
F;720-772/Domain: complement factor H repeat homology <FH12>
F;777-837/Domain: complement factor H repeat homology <FH13>
F;846-901/Domain: complement factor H repeat homology <FH14>
F;909-965/Domain: complement factor H repeat homology <FH15>
F;970-1026/Domain: complement factor H repeat homology <FH16>
F;1034-1056/Domain: transmembrane #status predicted <TMM>
F;1057-1091/Domain: intracellular #status predicted <INT>
F;121,127,294,372,612,698,858,881,919/Binding site: carbohydrate (Asn) (covalent) #st

  Query Match             62.0%;  Score 466.5;  DB 1;  Length 1091;
  Best Local Similarity   61.2%;  Pred. No. 1.4e-36;
  Matches   82;  Conservative   18;  Mismatches   48;  Indels    1;  Gaps    1;

Qy       2 ISCDPPPFVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           |||  || : | |  ||| ||  |||:||:|  ::|||||:: ||:::|  |||| |
Db     27 ISCGSPPSVLNGRISYYSIPSAVGTVIRYSCSGTFKLIGESSLFCIKDKVDGTWDKPAP 86

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHCDSVTFTCKANFTMKGSKIVWCQANEMWG 121
           ||  ||  ||  :||||||:  :||  |:|| ||||| || ||:| |:|:|||||| |||
Db     87 KCLYFNKVSSCTEPIVPGGYKIRGS TPYRHGDSVTFACKANFSMNGNKSVWCQANNMWG 146

Qy     122 PTALPVCESDFPLE 135
           || || | | ||||
Db     147 PTRLPTCVSVFPLE 160


RESULT 5
S53711
C4BP alpha chain precursor - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Sep-1989 #sequence_revision 10-Sep-1999 #text_change 10 Sep 1999
C;Accession: S53711
R;de Frutos, P.G.; Wahlbaeck, B.
Biochim. Biophys. Acta 1261, 285-289, 1995
A;Title: cDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence
A;Reference number: S53711; MUID:95226458
A;Accession: S53711
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-597 <DEF>
A;Cross-references: EMBL:Z35490
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;50-107/Domain: complement factor H repeat homology <FH1>
F;112-169/Domain: complement factor H repeat homology <FH2>
F;174-234/Domain: complement factor H repeat homology <FH3>
F;239-294/Domain: complement factor H repeat homology <FH4>
F;299-360/Domain: complement factor H repeat homology <FH5>
F;364-422/Domain: complement factor H repeat homology <FH6>
```

F;426-480/Domain: complement factor H repeat homology <FH7>
F;484-538/Domain: complement factor H repeat homology <FH8>

Query Match 40.9%; Score 242; DB 1; Length 597;
Best Local Similarity 31.6%; Pred. No. 2e-14;
Matches 44; Conservative 28; Mismatches 54; Indels 6; Gaps 3;

Qy 2 ISCDPPPEVKNAKKPYYSLPI VPCTVLRYYCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
Db 173 VKCKPPPDIRNGRHSGEENFYAY ---GFSVTYSCDPRFSLIGHASISCTVENETIGVWR 228

Qy 61 PICESVNKTISCSDPIVPGGFMNKGSKAIFKKHGDSVTFTCKANFTMKGSKTVWCQANNMW 120
Db 229 PSPPTCEK----ITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCKTVLRGSSVIHCIDD 284

Qy 121 GPTALPVCESD 131
Db 285 NPSP PVCESN 295

RESULT 5
NBHUC4
C4b binding protein alpha chain precursor - human
N;Alternate names: C4BP; proline rich protein
C;Species: Homo sapiens (man)
C;Date: 13-Aug-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A33568; S02372; A90426; A24182; A93134; S29492; A31785; I52244; A03210
R;Matsushita, T.; Okamura, S.; Aso, T.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 165, 138-144, 1989
A;Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity o
A;Reference number: A33568; MUID:90074699
A;Accession: A33568
A;Molecule type: mRNA
A;Residues: 1-597 <MA1>
A;Cross-references: GB:M31452; NID:g190501; PIDN:AAA36507.1; PID:g190502
A;Note: the authors translated the codon GGA for residue 492 as Glu
R;Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.
FEBS Lett. 232, 328-332, 1988
A;Title: Derivation of the sequence of the signal peptide in human C4b binding protein a
A;Reference number: S02372; MUID:88242821
A;Accession: S02372
A;Molecule type: mRNA
A;Residues: 17-81 <LI2>
A;Cross-references: EMBL:X07853
A;Note: although the sequence determined extends to residue 9 above, these authors assig
R;Chung, L.P.; Bentley, D.R.; Reid, K.B.M.
Biochem. J. 230, 133-141, 1985
A;Title: Molecular cloning and characterization of the cDNA coding for C4b binding proto
A;Reference number: A90426; MUID:86025405
A;Accession: A90426
A;Molecule type: mRNA
A;Residues: 80-597 <CH2>
A;Cross-references: GB:X02885; NID:g29564; PIDN:CAA26617.1; PID:g29565
A;Note: 92 Thr and 357 His were also found
R;Lintin, S.J.; Reid, K.B.M.
FEBS Lett. 204, 77-81, 1986
A;Title: Studies on the structure of the human C4b binding protein gene.
A;Reference number: A24182; MUID:86301119
A;Accession: A24182
A;Molecule type: DNA
A;Residues: 203-288 <LIN>
A;Cross-references: EMBL:X04284; EMBL:X04296
R;Rodriguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.
J. Exp. Med. 173, 1073-1082, 1991
A;Title: Structure of the gene coding for the alpha polypeptide chain of the human compl
A;Reference number: A43023; MUID:91217619
A;Contents: annotation; exon-intron boundaries
R;Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol. Immunol. 22, 427-435, 1985
A;Title: Amino acid sequence studies of human C4b binding protein: N-terminal sequence a
ced by cyanogen bromide treatment.
A;Reference number: A93134; MUID:85296001
A;Accession: A93134
A;Molecule type: protein
A;Residues: 49-81 <CH1>
A;Note: this paper reports amino-terminal sequences of the intact protein and of a nu
R;Bessinq, M.; Kanters, D.; Takeya, H.; van't Veer, C.; Hackeng, T.M.; Iwanaga, S.; B
FEBS Lett. 317, 228-232, 1993
A;Title: The region Ser(334)-Arg(356) of the alpha chain of human C4b binding protein
A;Reference number: S29492; MUID:93146164
A;Accession: S29492
A;Status: preliminary
A;Molecule type: protein
A;Residues: 381-404 <HES>
R;Suzuki, K.; Nishioka, J.
J. Biol. Chem. 263, 17034-17039, 1988
A;Title: Binding site for vitamin K-dependent protein S on complement C4b binding pro
A;Reference number: A31785; MUID:89034204
A;Accession: A31785
A;Molecule type: protein
A;Residues: 495-505,'X',507-510,'X',512-515 <SUZ>
A;Note: this peptide appears to bind protein S
R;Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983
A;Title: Visualization of human C4b-binding protein and its complexes with vitamin K
A;Reference number: A93950; MUID:83221615
A;Contents: annotation; electron microscopy; three-dimensional structure; ligand bind
R;Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 174, 222-227, 1991
A;Title: Genomic organization of the alpha chain of the human C4b binding protein gen
A;Reference number: I52244; MUID:91113199
A;Accession: I52244
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-597 <ASO>
A;Cross-references: GB:M62486; NID:g190498; PIDN:AAA36506.1; PID:g190500
C;Comment: C4BP controls the classical pathway of complement activation. It binds as
the degradation of the C4b2a complex (C3 convertase) by dissociating the complement
C;Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment
isulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. Amino
actors V and VIII.
C;Comment: The molecule has a central body supporting seven tentacles (alpha chains),
C;Genetics:
A;Gene: GDB:C4BPA
A;Cross-references: GDB:120568; OMIM:120830
A;Map position: 1q32-1q32
A;Introns: 48/1; 315/2; 343/2; 372/2; 397/1; 397/2; 463/2; 525/2; 582/2; 596/3
C;Complex: octamer of seven alpha chains and one beta chain
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein;
F;1-48/Domain: signal sequence #status predicted <SIG>
F;49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>
F;50-108/Domain: complement factor H repeat homology <FH1>
F;113-170/Domain: complement factor H repeat homology <FH2>
F;175-234/Domain: complement factor H repeat homology <FH3>
F;239-294/Domain: complement factor H repeat homology <FH4>
F;299-360/Domain: complement factor H repeat homology <FH5>
F;364-422/Domain: complement factor H repeat homology <FH6>
F;381-404/Region: complement C4b binding #status predicted
F;426-480/Domain: complement factor H repeat homology <FH7>
F;484-538/Domain: complement factor H repeat homology <FH8>
F;221,506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 40.0%; Score 225.5; DB 1; Length 597;
Best Local Similarity 31.8%; Pred. No. 8.8e-14;
Matches 42; Conservative 25; Mismatches 52; Indels 14; Gaps 4;

Qy 2 ISCDPPPEVKNAR----KPYYSLPIVPCTVLRYYCSPSYRLIGEKAIFCISENQVHATWD 57
Db 173 VKCKPPPDIRNGRHSGEENFYAY ---GFSVTYSCDPRFSLIGHASISCTVENETIGVWR 228

Qy 58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAIFKKHGDSVTFTCKANFTMKGSKTVWCQAN 117
Db 229 PSPPTCEK----ITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCKTVLRGSSVIHCIDD 284

Qy 118 EMWGPTALPVCE 129

```
                | | :  | ||
Db     285 SKWNPSP PACE 295


RESULT 7
NBMSC4
'4b binding protein alpha chain precursor - mouse
N;Alternate names: C4BP; proline-rich protein
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Jul-1997
C;Accession: A27117
R;Kristensen, T.; Ogata, R.T.; Chung, L.P.; Reid, K.B.M.; Tack, B.F.
Biochemistry 26, 4668-4674, 1987
A;Title: cDNA structure of murine C4b binding protein, a regulatory component of the ser
A;Reference number: A27117; MUID:88024997
A;Accession: A27117
A;Molecule type: mRNA
A;Residues: 1-469 <KRI>
A;Note: the authors translated the codon GCT for residue 25 as Val
C;Comment: C4BP controls the classical pathway of complement activation. It binds as a c
 the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement fr
C;Comment: In mouse, C4BP is a multimeric protein of noncovalently associated chains.
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla
F;1-56/Domain: signal sequence #status predicted <SIG>
F;57-469/Product: C4b-binding protein alpha chain #status predicted <MAT>
F;58-115/Domain: complement factor H repeat homology <FH1>
F;120-176/Domain: complement factor H repeat homology <FH2>
F;181-240/Domain: complement factor H repeat homology <FH3>
F;245-299/Domain: complement factor H repeat homology <FH4>
F;304-355/Domain: complement factor H repeat homology <FH5>
F;359-414/Domain: complement factor H repeat homology <FH6>
F;74,227,275,292,366,381,428/Binding site: carbohydrate (Asn) (covalent) #status predict

  Query Match             29.5%;  Score 222;  DB 1;  Length 469;
  Best Local Similarity   34.3%;  Pred. No. 1.4e-13;
  Matches   47;  Conservative   19;  Mismatches   57;  Indels   14;  Gaps    5;

Qy       2 ISCDPPPEVKNAR----KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           : | |||:: | :    : :|  |   |  : ||| | :||:|   | |   |:    |
Db     179 VKCGPPPDISNGKHSGTEDFY--PYNHG  ISYTCDPGFRLVGSPFIGCTVVNKTVPVWS 234

Qy      58 KAPPICFSVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             :|| ||     | || | :  | :  | || : | |||   |     ::|   : || |
Db     235 SSPPTCEK - --IICSQPNILHGVIVSGYKATYTHRDSVRLACLNGTVILRGRHVIECQGN 290

Qy     118 EMWGPTALPVCFSDFPI 134
             |  ::|| || |  |
Db     291 GNW- SSLPTCEFDCDI 305


RESULT 8
I36936
complement receptor 1 - chimpanzee (fragment)
C;Species: Pan troglodytes (chimpanzee)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I36936; I36937
R;Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J. Immunol. 153, 691-700, 1994
A;Title: Primary sequence of an alternatively spliced form of CR1. Candidate for the 75,
A;Reference number: I36935; MUID:94292799
A;Accession: I36936
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2014 <BIR1>
A;Cross-references: GB:L24920; NID:g557564; PIDN:AAA51438.1; PID:g557725
A;Accession: I36937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-397,1751-2014 <BIR2>
A;Cross-references: GB:L24921; NID:g557726; PIDN:AAA51439.1; PID:g557727
C;Genetics:
A;Gene: CR1
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: duplication; glycoprotein; tandem repeat
F;18-74/Domain: complement factor H repeat homology <FH01>
F;79-136/Domain: complement factor H repeat homology <FH02>
F;147-207/Domain: complement factor H repeat homology <FH03>
F;213-268/Domain: complement factor H repeat homology <FH04>
F;272-328/Domain: complement factor H repeat homology <FH05>
F;333-391/Domain: complement factor H repeat homology <FH06>
F;396-462/Domain: complement factor H repeat homology <FH07>
F;470-526/Domain: complement factor H repeat homology <FH08>
F;529-586/Domain: complement factor H repeat homology <FH09>
F;722-778/Domain: complement factor H repeat homology <FH10>
F;1041-1107/Domain: complement factor H repeat homology <FH11>
F;1172-1228/Domain: complement factor H repeat homology <FH12>
F;1233-1291/Domain: complement factor H repeat homology <FH13>
F;1296-1362/Domain: complement factor H repeat homology <FH14>
F;1432-1489/Domain: complement factor H repeat homology <FH15>
F;1625-1681/Domain: complement factor H repeat homology <FH16>
F;1749-1815/Domain: complement factor H repeat homology <FH17>
F;1823-1879/Domain: complement factor H repeat homology <FH18>
F;1884-1940/Domain: complement factor H repeat homology <FH19>

  Query Match             29.5%;  Score 221.5;  DB 2;  Length 2014;
  Best Local Similarity   36.8%;  Pred. No. 7.2e-13;
  Matches   50;  Conservative   17;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||      |||:   |          : |:|||:|:| |::
Db    1492 ISCEPPPFISNG  DFYSNNRASFHNGTVVIYQCHTGFDGFQLFELVGERSIYCTSKDDQ 1549

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRIGDSVTFTCKANFTMKGSKIV 112
             |   || | | ||   |: | |        |: : |   : | | |:  | | || ||
Db    1550 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSITEIVRFRCQPGFVMVGSHIV 1606

Qy     113 WCQANEMWGPTALPVC 128
           || |  ||   || |
Db    1607 QCQTNGRWGP-KLPHC 1621


RESULT 9
I73012
complement C3b/C4b receptor, membrane-bound form precursor - human
N;Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); sur
N;Contains: complement C3b/C4b receptor, secreted form
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 24-Nov-1999 #text_change 21-Jul-2000
C;Accession: I73012; I56203; A47602; S03291; S03843; A28587; A24748; B24748; C24748
R;Vik, D.P.; Wong, W.W.
J. Immunol. 151, 6214-6224, 1993
A;Title: Structure of the gene for the F allele of complement receptor type 1 and seq
A;Reference number: I56203; MUID:94065175
A;Accession: I73012
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-683,'X',685-1133,'X',1135-1471,'X',1473-2489 <VIK>
A;Cross-references: GB:L17418; NID:g306678; PIDN:AAB60695.1; PID:g451303
A;Accession: I56203
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-683,'X',685-894,'A',896-1000,1451-1471,'X',1473-2489 <VIK2>
A;Cross-references: GB:L17418; NID:g306678; PIDN:AAB60694.1; PID:g306680
R;Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.
J. Exp. Med. 169, 847-863, 1989
A;Title: Structure of the human CR1 gene. Molecular basis of the structural and quant
A;Reference number: A47602; MUID:89176869
A;Accession: A47602
A;Molecule type: DNA
A;Residues: 1-41 <WON>
A;Cross-references: GB:X14893
R;Hourcade, D., Miesner, D.R.; Atkinson, J.P., Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b r
 type 1.
```

A;Reference number: S03291; MUID:89001527
A;Accession: S03291
A;Molecule type: mRNA
A;Residues: 26-504 <HOD>
A;Cross-references: EMBL:X14942; NID:g30197; PIDN:CAA32541.1; PID:g736240
A;Experimental source: clone CR1.4
R;Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J. Exp. Med. 168, 1699-1717, 1988
A;Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b r
A;Reference number: S03843; MUID:89045992
A;Accession: S03843
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-891,'A',896-1000,1451-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2
A;Cross-references: EMBL:Y00816; NID:g30185; PIDN:CAA68755.1; PID:g30186
R;Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J. Exp. Med. 165, 1095-1112, 1987
A;Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domain
A;Reference number: A28507; MUID:87168191
A;Accession: A28507
A;Molecule type: mRNA
A;Residues: 953-1221,'TV',1224-2064,'I',2066-2276,'P',2278-2299,'H',2301-2325,'T',2327-2
A;Cross-references: GB:X05430; NID:g30196; PIDN:CAA28933.1; PID:g309019
R;Wong, W.W.; Klickstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A;Title: Identification of a partial cDNA clone for the human receptor for complement fr
A;Reference number: A94073; MUID:86067975
A;Accession: A24748
A;Molecule type: mRNA
A;Residues: 411-434,729-745,841-845 <WO2>
A;Cross-references: GB:M11564; NID:g180991; PIDN:AAA52297.1; PID:g180995; GB:M11617; NID
C;Genetics
A;Gene: GDB:CR1; CD35
A;Cross-references: GDB:119800; OMIM:120620
A;Map position: 1q42-1q42
A;Introns: 41/1; 1/1; 34/2; 78/3; 206/3; 65/1; 80/2; 418/1; 141/3; 551/1; 583/2; 6
; 1494/2; 114/1; 2146/3; 57/3; 37<e/1; 2708/2; 1943/3; 1941/1; 1<5/1; 2099/1
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-41/Domain: signal sequence #status predicted <SIG>
F;42-2489/Product: complement C3b/C4b receptor, membrane bound form #status predicted <M
F;42-1984/Product: complement C3b/C4b receptor, secreted form #status predicted <MAT2>
F;44-99/Domain: complement factor H repeat homology <FH01>
F;104-161/Domain: complement factor H repeat homology <FH02>
F;165-222/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;287-354/Domain: complement factor H repeat homology <FH05>
F;358-416/Domain: complement factor H repeat homology <FH06>
F;421-487/Domain: complement factor H repeat homology <FH07>
F;493-549/Domain: complement factor H repeat homology <FH08>
F;554-611/Domain: complement factor H repeat homology <FH09>
F;615-682/Domain: complement factor H repeat homology <FH10>
F;688-744/Domain: complement factor H repeat homology <FH11>
F;747-803/Domain: complement factor H repeat homology <FH12>
F;808-866/Domain: complement factor H repeat homology <FH13>
F;871-937/Domain: complement factor H repeat homology <FH14>
F;963-999/Domain: complement factor H repeat homology <FH15>
F;1004-1061/Domain: complement factor H repeat homology <FH16>
F;1066-1132/Domain: complement factor H repeat homology <FH17>
F;1138-1193/Domain: complement factor H repeat homology <FH18>
F;1197-1253/Domain: complement factor H repeat homology <FH19>
F;1258-1316/Domain: complement factor H repeat homology <FH20>
F;1321-1387/Domain: complement factor H repeat homology <FH21>
F;1393-1449/Domain: complement factor H repeat homology <FH22>
F;1454-1511/Domain: complement factor H repeat homology <FH24>
F;1516-1582/Domain: complement factor H repeat homology <FH24>
F;1588-1643/Domain: complement factor H repeat homology <FH25>
F;1647-1703/Domain: complement factor H repeat homology <FH26>
F;1708-1766/Domain: complement factor H repeat homology <FH27>
F;1771-1837/Domain: complement factor H repeat homology <FH28>
F;1846-1902/Domain: complement factor H repeat homology <FH29>
F;1907-1964/Domain: complement factor H repeat homology <FH30>
F;1969-2035/Domain: complement factor H repeat homology <FH31>
F;2041-2096/Domain: complement factor H repeat homology <FH32>
F;2100-2156/Domain: complement factor H repeat homology <FH33>
F;2161-2219/Domain: complement factor H repeat homology <FH34>
F;2224-2290/Domain: complement factor H repeat homology <FH35>
F;2298-2354/Domain: complement factor H repeat homology <FH36>
F;2359-2415/Domain: complement factor H repeat homology <FH37>

Query Match 29.5%; Score 221.5; DB 2; Length 2489;
Best Local Similarity 36.8%; Pred. No. 9e-13;
Matches 50; Conservative 17; Mismatches 54; Indels 15; Gaps 5;

```
Qy       2 ISCTPPPEVENARRFYYS---LRYVPGIVLRYSGS       YVLLSEKATFLYGVNQV 52
           |||:|||  : |    :||         |||: | |         :  :|||:|| |::
Db    1967 LSCEPPPTIPNGNG  DFYSNNRTSFHNGTVVTYQCHTGFVGSDEVICVESGSWTSLPAEDQ 2024

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGPMNKSKAPTFRHGDSVTFTCKANFTMKGSKTV 112
             |  || | | ||   |: | |        ::: |  : |  :  | | | || ||
Db    2025 VGVWSSPPPRCISTNK  CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 2081

Qy     113 WCQANRMWGPTALPVC 128
             || |  ||| || |
Db    2082 QCQTNGRWGP KLPHC 2096


RESULT 10
A56740
sperm-egg recognition protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 20-Aug-1999
C;Accession: A56740
R;Bookbinder, L.H.; Cheng, A.; Bleil, J.D.
Science 269, 86-89, 1995
A;Title: Tissue- and species-specific expression of sp56, a mouse sperm fertilization
A;Reference number: A56740; MUID:95327956
A;Accession: A56740
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-579 <HOO>
A;Cross-references: GB:U17108; NID:g897562; PIDN:AAC52208.1; PID:g897563
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
C;Keywords: fertilization; homomultimer; peripheral membrane protein; sperm
F;1-32/Domain: signal sequence #status predicted <SIG>
F;34-91/Domain: complement factor H repeat homology <FH1>
F;96-153/Domain: complement factor H repeat homology <FH2>
F;158-218/Domain: complement factor H repeat homology <FH3>
F;223-278/Domain: complement factor H repeat homology <FH4>
F;283-346/Domain: complement factor H repeat homology <FH5>
F;349-411/Domain: complement factor H repeat homology <FH6>
F;456-509/Domain: complement factor H repeat homology <FH7>

Query Match 28.6%; Score 215; DB 2; Length 579;
Best Local Similarity 30.3%; Pred. No. 8.1e-13;
Matches 40; Conservative 27; Mismatches 51; Indels 14; Gaps 4;

Qy       5 SCPDPGFEVENARPKDYYSIPIVPGIVLRYVSPSYRLTGSEKATFLYSENQVHAIWD 62
           :|:||| : | :        |:    |:|: | | |:: |:|  :| |   |:    |
Db     157 TCEPPPPISNGKHSGRDDDLYTF----GSVVIYNCDPTFTLLGNASIVCTVVNKTVSWR 212

Qy      63 KAPPICESVNKTISCSDPIVPGGPMNKGSKAPTFRHGDSVTFTCKANFTMKGSKTVWTQAN 122
             ||  |  | |  | |:|:|::  | :  : : |::   ||    :|||  : :|||
Db     213 PHPPACQKR---IVCHRPVIPKGYLAPGFRQEYAYRDALDIRCKNGFTLRGSSVIHTEAN 268

Qy     123 EMWGPTALPVCE 134
             | | ::| ||
Db     269 GEWFPSIPTCE 280


RESULT 11
S57953
C4BP protein alpha chain precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
C;Accession: S57953
```

```
R;Hillarp, A.; Thern, A.; Dahlback, B.
submitted to the EMBL Data Library, July 1995
A;Description: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: stru
A;Reference number: S57953
A;Accession: S57953
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-558 <HIL>
A;Cross-references: EMBL:Z50051; NID:g899379; PIDN:CAA90391.1; PID:g899380
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;15-72/Domain: complement factor H repeat homology <FH1>
F;77-134/Domain: complement factor H repeat homology <FH2>
F;139-199/Domain: complement factor H repeat homology <FH3>
F;204-258/Domain: complement factor H repeat homology <FH4>
F;263-324/Domain: complement factor H repeat homology <FH5>
F;328-386/Domain: complement factor H repeat homology <FH6>
F;390-443/Domain: complement factor H repeat homology <FH7>
F;447-501/Domain: complement factor H repeat homology <FH8>

  Query Match              27.7%;  Score 208.5;  DB 2;  Length 558;
  Best Local Similarity    30.9%;  Pred. No. 3.2e-12;
  Matches   43;  Conservative   26;  Mismatches   51;  Indels   19;  Gaps    5;

Qy       4 CDPPPEVKNAR---- KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDK 58
           |  ||:: | :     : :::     : : | | | : |:|  :|  |:    |
Db     139 CGMPPDISNGKHNGREEEFFYY  - RSSVIYKCDPDFILLGNASITCIVVNKTVGVWSP 194

Qy      59 APPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQANE 118
           :|| ||     | |  | |  | :| | |  ::: ||| | |:  | ::||  : |:|:
Db     195 SPPICE    RICPWPKVIHGTINSGFKGTLYKYKDSVRFVCQKGFVIRGSGVIHCEADG 250

Qy     119 MWGPTALPVCE- --SDFP 133
           | |  :||||    :| |
Db     251 SWSP  VPVCELNSCTDIP 267


RESULT 12
146001
C4b binding protein alpha chain  bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: 146001; S43190
R;Hillarp, A.; Thern, A.; Dahlback, B.
J. Immunol. 153, 4190-4199, 1994
A;Title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains pro
A;Reference number: 146001; MUID:95015909
A;Accession: 146001
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-610 <HIL>
A;Cross-references: EMBL:Z31693; NID:g469117; PIDN:CAA83498.1; PID:g469118
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
F;50-107/Domain: complement factor H repeat homology <FH1>
F;112-169/Domain: complement factor H repeat homology <FH2>
F;174-234/Domain: complement factor H repeat homology <FH3>
F;239-294/Domain: complement factor H repeat homology <FH4>
F;299-362/Domain: complement factor H repeat homology <FH5>
F;366-425/Domain: complement factor H repeat homology <FH6>
F;429-483/Domain: complement factor H repeat homology <FH7>
F;487-541/Domain: complement factor H repeat homology <FH8>

  Query Match              26.6%;  Score 200;  DB 1;  Length 610;
  Best Local Similarity    27.5%;  Pred. No. 2.3e-11;
  Matches   46;  Conservative   27;  Mismatches   54;  Indels   14;  Gaps    4;

Qy       4 CDPPPEVKNAR-- - KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDK 58
           |:||| : | |     : :::     |: : |:|    : :::: :| |  ||:    |
Db     174 CEPPPPIISNGRHNGGDEDEYIY-- -GSSVIYSCDRDFSMLGKASISCRVENKTIGVWSP 229

Qy      59 APPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVIFICKANFIMKGSKTVWCQANE 118
           :|| |     | : |  |:|  | :  |     : :  |: : |    | ::|    : |:|:
Db     230 SPPSC  - KKVICVQPVVKDGKIISGFGPIYIYQQSIVYACNKGFRLEGDSLIHCEADN 285

Qy     119 MWGPTALPVCE 129
           | |  | |
Db     286 SWNPPP-PICE 295


RESULT 13
C36838
complement control protein homolog - variola virus (strain India 1967)
N;Alternate names: D12L protein
C;Species: variola virus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Mar-2000
C;Accession: C36838
R;Blinov, V.M.
submitted to GenBank  November 1992
A;Reference number: A36859
A;Accession: C36838
A;Molecule type: DNA
A;Residues: 1-263 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA48953.1; PID:g297195
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
F;21-81/Domain: complement factor H repeat homology <FH1>
F;86-143/Domain: complement factor H repeat homology <FH2>
F;148-201/Domain: complement factor H repeat homology <FH3>
F;206-261/Domain: complement factor H repeat homology <FH4>

  Query Match              26.3%;  Score 197.5;  DB 1;  Length 263;
  Best Local Similarity    29.9%;  Pred. No. 1.1e-11;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |:     |:|: |:|:  :| |||   : |     |   ||
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSN PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           |:     : |  | :  |::: | |  : : |:| |||  : : || :  |     |
Db     200 TCQ----IVKCPHPTILNGYLSSGFKRSYSYNDNVDFICKYGYKLSGSSSSICSPGNTWQ 255

Qy     122 PTALPVC 128
           |  | |
Db     256 P-ELPKC 261


RESULT 14
T28450
hypothetical protein D15L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28450
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox vir
A;Reference number: Z20488; MUID:94088747
A;Accession: T28450
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-263 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60760.1; PID:g438930
A;Experimental source: strain "Bangladesh-1975"
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom

  Query Match              26.3%;  Score 197.5;  DB 2;  Length 263;
  Best Local Similarity    29.9%;  Pred. No. 1.7e-11;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |:     |:|: |:|:  :| |||   : |     |   ||
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSN PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVIFICKANFIMKGSKTVWCQANEMWG 121
           |:     :   | :  |::: | |  : : |:| |||  : : || :  |     |
Db     206 TCQ----IVKCPHPTILNGYLSSGFKRSYSYNDNVDFICKYGYKLSGSSSSICSPGNTWQ 255

Qy     122 PTALPVC 128
```

```
                ; ||| |
Db     256 P-ELPKC 261


RESULT 15
B72152
B18L protein - variola minor virus (strain Garcia 1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: B72152
R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Lopar
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of classic variola minor
A;Reference number: A72150
A;Accession: B72152
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <SHC>
A;Cross-references: GB:AY186780; NID:g5830555; PIDN:AAB54612.1; PID:g5830573
A;Experimental source: strain Garcia 1966
C;Genetics:
A;Gene: B18L
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo

  Query Match             26.0%;  Score 195.5;  DB 2;  Length 263;
  Best Local Similarity   29.9%;  Pred. No. 2.6e-11;
  Matches   58;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           |  || : | |   |:      : :|: |:|:  | |||   : |          |   ||
Db     146 VKCQLPPSISNGKRHNGYNDFYTDGSVVLYSCNSGYSLIGNSGVLCSG     GEWSN-PP 199

Qy      62 TCESVNKTISCSDPIVPGGFMNKGSKAPFFHDGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           |:      | | :  |:: | |  : : |:| |||  : : || :  |      |
Db     200 TCQ     IVKCPYPFTILNGYLSSGFKRSYSYNDNVDFTCKYGYKLSGSSSSFCSPGNTWQ 255

Qy     122 PTALPVC 128
           |  || |
Db     256 P-ELPKC 261


RESULT 16
WMVZSP
apolipoprotein H homolog precursor - vaccinia virus
N;Alternate names: 35K secretory protein; C3L protein; virokine
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1989 #sequence_revision 30-Jun-1990 #text_change 22-Jun-1999
C;Accession: A31005; B42504
R;Kotwal, G.J.; Moss, B.
Nature 335, 176-178, 1988
A;Title: Vaccinia virus encodes a secretory polypeptide structurally related to complem
A;Reference number: A31005; MUID:88318974
A;Accession: A31005
A;Molecule type: DNA
A;Residues: 1-263 <KOT>
A;Cross-references: GB:X15166; NID:g60690; PIDN:CAA33164.1; PID:g60691
A;Experimental source: strain WR
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A;Title: Appendix to "The complete DNA sequence of vaccinia virus".
A;Reference number: A42501
A;Accession: B42504
A;Molecule type: DNA
A;Residues: 1-263 <GOE>
A;Cross-references: GB:M35027; NID:g335811; PIDN:AAA47997.1; PID:g335845
A;Experimental source: strain Copenhagen
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A;Title: The complete DNA sequence of vaccinia virus.
A;Reference number: A42581; MUID:91021027
A;Contents: annotation; possible protein coding frames
A;Note: neither amino acid nor nucleotide sequence is given

C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C;Keywords: duplication; extracellular protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-263/Product: C4b binding protein homolog #status predicted <MAT>
F;21-81/Domain: complement factor H repeat homology <FH1>
F;86-143/Domain: complement factor H repeat homology <FH2>
F;148-201/Domain: complement factor H repeat homology <FH3>
F;206-261/Domain: complement factor H repeat homology <FH4>

  Query Match             25.7%;  Score 194.5;  DB 1;  Length 263;
  Best Local Similarity   29.1%;  Pred. No. 4e-11;
  Matches   47;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSIPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |       |:|: |:|:  | |||   : |          |   ||
Db     146 VKCQSPPSISNGKRHNGYEDFYTDGSVVTYSCNSGYSLIGNSGVLCSG     GEWSN-PP 199

Qy      62 TCESVNKTISCSDPIVPGGFMNKGSKAPFFHDGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           |:      | | :  |:: | |  : : |:| |||  : : || :  |      |
Db     200 TCQ     IVKCPYPFTILNGYLSSGFKRSYSYNDNVDFTCKYGYKLSGSSSSFCSPGNTWQ 255

Qy     122 PTALPVC 128
           |  || |
Db     256 P-ELPKC 261


RESULT 17
G02913
sperm CD46 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000
C;Accession: G02913
R;Hara, T.
submitted to GenBank, March 1996
A;Reference number: H01942
A;Accession: G02913
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-349 <HAR>
A;Cross-references: GB:D84105; NID:g1256700; PIDN:BAA12224.1; PID:g1256701
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 349;
  Best Local Similarity   30.6%;  Pred. No. 1.6e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS   -  YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : :| | :     : |||  :|  :| |
Db     159 KVLCIPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV - 215

Qy      55 TWDKAPPTCESVNKTISCSDPIVPGGFMNKGSKAEFHHGDSVTFTCKANFTMKGSKTVWC 114
             | :|   :    | : |  |:|  |    |    | :  :| | |   | : |: : |
Db     216 -WSRAAPEC-  - KVVKCRFPVVENGKQISGFGKKFYYKAIVMFECDKGFYLEGSDIIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  : :|  :| |
Db     271 DSNSTWDP PVPKC 283


RESULT 18
I57998
membrane cofactor protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 12-May-2000
C;Accession: I57998
R;Cervoni, F.; Fenichel, P.; Akhoundi, C.; Hsi, B.L.; Rossi, B
Mol. Reprod. Dev. 34, 107-113, 1993
A;Title: Characterization of a cDNA clone coding for human testis membrane cofactor p
A;Reference number: I57998; MUID:93119658
```

```
A;Accession: I57998
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-369 <RES>
A;Cross-references: GB:S51940; NID:g262937; PIDN:AAB24802.1; PID:g262938
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 369;
  Best Local Similarity   30.6%;  Pred. No. 1.7e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVIRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 IWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFEHGDSVIFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216 -WSRAAPEC  - KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 19
I54479
membrane cofactor protein precursor, splice form pm5.1 - human
N;Alternate names: lymphocyte surface glycoprotein CD46
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
C;Accession: I54479
R;Purcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.F.
Immunogenetics 44, 345-344, 1991
A;Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator o
A;Reference number: I54479; MUID:91267562
A;Accession: I54479
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-377 <RES>
A;Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
C;Genetics:
A;Gene: GDB:MCP
A;Cross-references: GDB:120169; OMIM:120920
A;Map position: 1q32-1q32
C;Function:
A;Description: for the factor I mediated cleavage of the complement convertases
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: alternative splicing; glycoprotein; transmembrane protein
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;329-351/Domain: transmembrane #status predicted <TMM>

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 377;
  Best Local Similarity   30.6%;  Pred. No. 1.8e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVIRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 IWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFEHGDSVIFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216  WSRAAPEC---KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 20
S01896
membrane cofactor protein precursor - human
N;Alternate names: lymphocyte surface glycoprotein CD46
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 21-Jul-2000
C;Accession: S01896; A60765; I56188
R;Lublin, D.M.; Liszewski, M.K.; Post, T.W.; Arce, M.A.; Le Beau, M.M.; Rebentisch, M
J. Exp. Med. 168, 181-194, 1988
A;Title: Molecular cloning and chromosomal localization of human membrane cofactor pr
A;Reference number: S01896; MUID:88286080
A;Accession: S01896
A;Molecule type: mRNA
A;Residues: 1-384 <LUB>
A;Cross-references: EMBL:Y00651; NID:g34504; PIDN:CAA68675.1; PID:g34505
A;Note: part of this sequence, including the amino end of the mature protein, was con
R;Purcell, D.F.J.; Deacon, N.J.; Andrew, S.M.; McKenzie, I.F.C.
Immunogenetics 31, 21-28, 1990
A;Title: Human non-lineage antigen, CD46 (LuLy-m5): purification and partial sequenci
A;Reference number: A60765; MUID:90129152
A;Accession: A60765
A;Molecule type: protein
A;Residues: 'X',36-39,'X',41-58 <PUR>
R;Cui, W.; Bourcade, D.; Post, T.; Greenwood, A.C.; Atkinson, J.P.; Kumar, V
J. Immunol. 151, 4137-4146, 1993
A;Title: Characterization of the promoter region of the membrane cofactor protein (CD
A;Reference number: I56188; MUID:94014356
A;Accession: I56188
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-94 <RES>
A;Cross-references: GB:S65879; NID:g425644; PIDN:AAC13968.1; PID:g1261968
C;Genetics:
A;Gene: GDB:MCP
A;Cross-references: GDB:120169; OMIM:120920
A;Map position: 1q32-1q32
C;Function:
A;Description: membrane cofactor protein for the factor I-mediated cleavage of the co
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C;Keywords: alternative splicing; glycoprotein; transmembrane protein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-384/Product: membrane cofactor #status experimental <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-157/Domain: complement factor H repeat homology <FH02>
F;162-223/Domain: complement factor H repeat homology <FH03>
F;228-283/Domain: complement factor H repeat homology <FH04>
F;329-351/Domain: transmembrane #status predicted <TMM>

  Query Match             25.1%;  Score 188.5;  DB 2;  Length 384;
  Best Local Similarity   30.6%;  Pred. No. 1.8e-10;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVIRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :: | |||::|| :  :  : :       : |:| |:      : ||||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC-GDNSV-- 215

Qy      55 IWDKAPPICESVNKLISCSDPIVPGGPMNKGSKAPFEHGDSVIFTCKANFTMKGSKTVWC 114
             | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216  WSRAAPEC---KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|  | |  :| |
Db     271 DSNSTWDP-PVPKC 283


RESULT 21
JC5194
membrane cofactor protein precursor - green monkey
N;Alternate names: CD46; MCP; measles virus receptor
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 19-May-2000
```

```
C;Accession: JC5194
R;Murakami, Y.; Soya, T.; Kurita, M.; Nagasawa, S.
Biol. Pharm. Bull. 19, 1541-1545, 1996
A;Title: Molecular cloning of a complementary DNA for a membrane cofactor protein (MCP,
A;Reference number: JC5148; MUID:97149841
A;Contents: Vero cell
A;Accession: JC5194
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-362 <MUR>
A;Cross-references: DDBJ:D78568
C;Comment: This protein is a complement regulator and measles virus receptor. It protect
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
C;Keywords: glycoprotein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-362/Product: membrane cofactor protein #status predicted <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-152/Domain: complement factor H repeat homology <FH02>
F;162-220/Domain: complement factor H repeat homology <FH03>
F;220-283/Domain: complement factor H repeat homology <FH04>
F;315-336/Domain: transmembrane #status predicted <TMM>
F;337-362/Domain: intracellular #status predicted <INT>
F;86,114,212,,13,263,273/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              24.1%;  Score 181.5;  DB 2;  Length 362;
  Best Local Similarity    29.1%;  Pred. No. 7.8e-10;
  Matches   39;  Conservative   25;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG TVLRYTCSPS      YRLIGEKAIFCISENQVHA 54
           |  ||| ||  :  :  :  :  :  : |:| |:      : |||| |:| :    ::
Db     159 KILCIPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN    NS 214

Qy      55 IWDFAFFCTECSVNKTISCSSPLVPGCFMNRSQSKAFERHGIIVTEICKANPWKKEIVW 114
           ||  | | |    |  :| |:|  |    |  :  :| | |    :  :|  :  |
Db     215 IWSHAAPEC    KVVKCREPVVENGKQISGFGKKFYYKA VMEECDKGYYLNGSDKIVC 270

Qy     115 QANEMWGPTALPVC 128
           ::|  | |  :| |
Db     271 ESNSIWDP PVPKC 283


RESULT 22
JC5148
membrane cofactor protein (MCP, CD46) receptor precursor - green monkey
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 01-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C;Accession: JC5148
R;Murakami, Y.; Soya, T.; Kurita, M.; Nagasawa, S.
Biol. Pharm. Bull. 19, 1541-1545, 1996
A;Title: Molecular cloning of a complementary DNA for a membrane cofactor protein (MCP,
A;Reference number: JC5148; MUID:97149841
A;Contents: Vero cell
A;Accession: JC5148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-369 <MUR>
A;Cross-references: DDBJ:D03811; NID:g1777414; PID:g1777415
C;Comment: This protein is a complement regulator and measles virus receptor. It protect
C;Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-369/Product: membrane cofactor protein (MCP, CD46) receptor #status predicted <MAT>
F;35-94/Domain: complement factor H repeat homology <FH01>
F;99-152/Domain: complement factor H repeat homology <FH02>
F;162-220/Domain: complement factor H repeat homology <FH03>
F;220-283/Domain: complement factor H repeat homology <FH04>
F;315-336/Domain: transmembrane #status predicted <TMM>
F;337-369/Domain: intracellular #status predicted <INT>

  Query Match              24.1%;  Score 181.5;  DB 2;  Length 369;
  Best Local Similarity    29.1%;  Pred. No. 8e-10;
  Matches   39;  Conservative   25;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVENARKPYYSLPIVPG TVLRYTCSPS      YRLIGEKAIFCISENQVHA 54
           |  ||| ||  :  :  :  :  :  : |:| |:      : |||| |:| :    ::
Db     159 KILCIPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN    NS 214

Qy      55 IWDFAFFCTECSVNKTISCSSPLVPGCFMNRSQSKAFERHGIIVTEICKANPWKKEIVW 114
           ||  | | |    |  :| |:|  |    |  :  :| | |    :  :|  :  |
Db     215 IWSHAAPEC----KVVKCREPVVENGKQISGFGKKFYYKAVMEECDKGYYLNGSDKIVC 270

Qy     115 QANEMWGPTALPVC 128
           ::|  | |  :| |
Db     271 ESNSIWDP PVPKC 283


RESULT 23
NBMSH
complement factor H precursor - mouse
N;Alternate names: protein beta-1-H
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999
C;Accession: A26154; I49711; I49728
R;Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A;Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt
A;Reference number: A26154; MUID:86233453
A;Accession: A26154
A;Molecule type: mRNA
A;Residues: 1-1234 <KRI>
A;Cross-references: GB:M12660; NID:g194724; PIDN:AAA37759.1; PID:g387181
R;Natsuume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A;Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
A;Reference number: I49711; MUID:90111033
A;Accession: I49711
A;Status: preliminary; translated from GB/DDBJ
A;Molecule type: mRNA
A;Residues: 1-18 <RES>
A;Cross-references: GB:M31979; NID:g194726; PIDN:AAA37762.1; PID:g194729
R;Munoz-Canoves, P.; Tack, B.F.; Vik, D.P.
Biochemistry 28, 9891-9897, 1989
A;Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
A;Reference number: I49728; MUID:90148935
A;Accession: I49728
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-19 <RE2>
A;Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926
C;Comment: Two codominant alleles of factor H are present in mice.
C;Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot
 (C5 convertase) in the alternative complement pathway.
C;Genetics:
A;Map position: 1
C;Superfamily: complement factor H; complement factor H repeat homology
C;Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1234/Product: complement factor H #status predicted <MPT>
F;21-80/Domain: complement factor H repeat homology <FH01>
F;85-141/Domain: complement factor H repeat homology <FH02>
F;146-205/Domain: complement factor H repeat homology <FH03>
F;210-262/Domain: complement factor H repeat homology <FH04>
F;246-248/Region: cell attachment (R-G-D) motif
F;267-320/Domain: complement factor H repeat homology <FH05>
F;325-385/Domain: complement factor H repeat homology <FH06>
F;389-442/Domain: complement factor H repeat homology <FH07>
F;448-505/Domain: complement factor H repeat homology <FH08>
F;509-564/Domain: complement factor H repeat homology <FH09>
F;569-622/Domain: complement factor H repeat homology <FH10>
F;629-683/Domain: complement factor H repeat homology <FH11>
F;690-743/Domain: complement factor H repeat homology <FH12>
F;752-802/Domain: complement factor H repeat homology <FH13>
F;808-861/Domain: complement factor H repeat homology <FH14>
F;867-921/Domain: complement factor H repeat homology <FH15>
F;936-989/Domain: complement factor H repeat homology <FH16>
F;994-1048/Domain: complement factor H repeat homology <FH17>
F;1053-1107/Domain: complement factor H repeat homology <FH18>
```

```
F;1114 1168/Domain: complement factor H repeat homology <FH19>
F;1172 1233/Domain: complement factor H repeat homology <FH20>
F;21 66,72 80,85 129,114 141,146 192,178 205,210 251,237 262,267 309,294 320,325 374,357
0 802,808 850,846 861,867 920,905 931,946 978,964 989,994 1037,1023 1048,1053 1096,1082
F;676,721,773,801,1040,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status pre

  Query Match             23.4%;  Score 176;  DB 1;  Length 1234;
  Best Local Similarity   37.4%;  Pred. No. 9.6e-09;
  Matches   52;  Conservative   15;  Mismatches   50;  Indels   22;  Gaps    9;

Qy        3 SCDPPPEVKN--ARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
            || || |: |    |:|  ||     | |::| | | : | |   | |:  |     |
Db      628 SCAPPLELLNGEINGAKKVEYS----HGEVVKYDCKPRFLLKGPNKIQCVDGN-----WT 678

Qy       58 KAPPICESVNKTISCSD-PIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQA 116
             |:|   :|  | | | :  | | |  |: ||||| | |: |||| |  :| | :
Db      679 TL PVCIFEERT--CGDIPELEHG-SAKCSVPPYHHGDSVEFICEENFTMIGHGSVSCIS 734

Qy      117 NEMWGPTALPVCESDFPLE 135
              :  | || | :   ||
Db      735 GK W  TQLPKCVATDQLE 750


RESULT 24
JC2054
complement regulatory protein, 512 antigen precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C;Accession: JC2054; PC2027
R;Sakurada, C.; Seno, H.; Doki, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.
Biochem. Biophys. Res. Commun. 198, 819-826, 1994
A;Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.
A;Reference number: JC2054; MUID:94161746
A;Accession: JC2054
A;Molecule type: mRNA
A;Residues: 1-497 <SAK>
A;Cross-references: GB:D42114; NID:g577641; PIDN:BAA07698.1; PID:g987688
A;Accession: PC2027
A;Molecule type: protein
A;Residues: 39-51 <SA2>
C;Comment: This protein plays a critical role in protection against complement mediated
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C;Keywords: glycoprotein; transmembrane protein
F;1-38/Domain: signal sequence #status predicted <SIG>
F;37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>
F;38-94/Domain: complement factor H repeat homology <FH1>
F;99-156/Domain: complement factor H repeat homology <FH2>
F;161-227/Domain: complement factor H repeat homology <FH3>
F;233-288/Domain: complement factor H repeat homology <FH4>
F;294-352/Domain: complement factor H repeat homology <FH5>
F;357-413/Domain: complement factor H repeat homology <FH6>
F;421-444/Domain: transmembrane #status predicted <TMM>
F;247,331,346,450,482,484/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             22.9%;  Score 172.5;  DB 2;  Length 497;
  Best Local Similarity   30.2%;  Pred. No. 7.9e-09;
  Matches   42;  Conservative   21;  Mismatches   55;  Indels   21;  Gaps    5;

Qy        2 ISCDPPPEVKNA-    RKPYYSLPIVPGTVLRYTCSPS- ----YRLIGEKAIFCISE 49
            | |: || : |      |: ::      | |: | |:         : |:|| :| | |
Db      159 IPCEIPPSIPNGDFFSPNREDFHY---GMVVTYQCNTDARGKKLFNLVGEPSIHCTSI 213

Qy       50 NQVHATWDKAPPICFSVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGS 109
            :     | || |  :||   |: | |     :   :|: |   | | | |:  | |||
Db      214 DGQVGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVEFRCQDGFMMKGD 270

Qy      110 KTVWCQANEMWGPTALPVC 128
              :|:|::   | |  || |
Db      271 SSVYCRSLNRWEP-QLPSC 288


RESULT 25
A43059
complement receptor CR1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C;Accession: D35069; A35070; B35070; C35070; D35070; E35069; E35070; F35069; F35070;
R;Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A;Title: Identification and sequence analysis of four complement factor H related tra
A;Reference number: A35070; MUID:90153969
A;Accession: D35069
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-808 <V K>
A;Cross-references: GB:M29009; NID:g192563; PIDN:AAA37416.1; PID:g309166; GB:J05259
A;Note: translation of the nucleotide sequence is not complete
C;Superfamily: complement factor H; complement factor H repeat homology
F;28-88/Domain: complement factor H repeat homology <FH01>
F;92-145/Domain: complement factor H repeat homology <FH02>
F;151-208/Domain: complement factor H repeat homology <FH03>
F;212-267/Domain: complement factor H repeat homology <FH04>
F;272-325/Domain: complement factor H repeat homology <FH05>
F;332-386/Domain: complement factor H repeat homology <FH06>
F;393-446/Domain: complement factor H repeat homology <FH07>
F;455-505/Domain: complement factor H repeat homology <FH08>
F;511-564/Domain: complement factor H repeat homology <FH09>
F;569-622/Domain: complement factor H repeat homology <FH10>
F;627-681/Domain: complement factor H repeat homology <FH11>
F;688-742/Domain: complement factor H repeat homology <FH12>
F;746-807/Domain: complement factor H repeat homology <FH13>

  Query Match             22.3%;  Score 168;  DB 2;  Length 808;
  Best Local Similarity   36.7%;  Pred. No. 3.9e-08;
  Matches   51;  Conservative   14;  Mismatches   52;  Indels   22;  Gaps    9;

Qy        3 SCDPPPEVKN-----ARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
            || |||: |      |:: ||     | |: | | | : |     | |     | |
Db      331 SCAQPPEIPNCEINGAKRVEYS----HGEVVGYDCKPRFLLKGPNKIQC----VDGMWT 381

Qy       58 KAPPICESVNKTISCSD-PIVPGGFMNKGSKA-FFHGDSVTFTCKANFTMKGSKTVWCQA 116
              |:|    :|  | | | :  | : | |  |:  |||| |||   ||| |  :| | :
Db      382 TL-PVCVEERT--CGDIPELEHGSV-KFSVP-YYHYGDSVEFICAEIFIMIGLGSVSCLS 437

Qy      117 NEMWGPTALPVCESDFPLE 135
              :  | .| | :   ||
Db      438 GK-W--TQLPKCVATDQLE 453


RESULT 26
A43519
complement receptor CR1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-Oct-1992 #sequence_revision 30-Jan-1993 #text_change 17-Nov 2000
C;Accession: A43519; A30550
R;Paul, M.S.; Aegerter, M.; Cepek, K.; Miller, M.D.; Weis, J.H.
J. Immunol. 144, 1988-1996, 1990
A;Title: The murine complement receptor gene family. The genomic and transcriptional
A;Reference number: A43519; MUID:90171600
A;Accession: A43519
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-440 <PAU>
A;Cross-references: GB:M34164
A;Note: the authors translated the codon GCC for residue 21 as Ala, and CAC for resid
R;Paul, M.S.; Aegerter, M.; O'Brien, S.E.; Kurtz, C.B.; Weis, J.H.
J. Immunol. 142, 582-589, 1989
A;Title: The murine complement receptor gene family. Analysis of mCRY gene products a
A;Reference number: A30550; MUID:89093944
A;Accession: A30550
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-20,'A',22-120,'E',122-433 <PAU>
C;Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
F;42-98/Domain: complement factor H repeat homology <FH1>
F;103-160/Domain: complement factor H repeat homology <FH2>
```

F;165-241/Domain: complement factor H repeat homology ·FH3·
F;247-294/Domain: complement factor H repeat homology ·FH4·
F;299-356/Domain: complement factor H repeat homology ·FH5·

Query Match 22.2%; Score 167; DB 2; Length 440;
Best Local Similarity 40.0%; Pred. No. 2.3e-08;
Matches 42; Conservative 21; Mismatches 55; Indels 22; Gaps 6;

Qy 3 ISCDPPPEVKNA RKRYYSLPIVPGTVLRYTCSPSY--YRLIGEKAIPCISEN 49
Db 163 IDCEIPPGIIPNGDFESSIREDFRY GMVVTYRCNTDARGKALFNLVGEPSLYCISN 217

Qy 50 NQVHAIWDKAPPICESVNKPISCS DPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKG 108
Db 218 DGEIGVWSSPPPQCIELNK CPPPYVENAVMLSENRSLFSLRDIVEFRCHPGFIMKG 274

Qy 109 SKIVWCQANEMWGPTALPVC 128
Db 275 ASSVRCQSLNKWEP ELPSC 293

RESULT 27
A34924
Complement C4b/C4b receptor like protein precursor - human
C;Species: Homo sapiens (man)
C;Date: 27-Jul-1990 #sequence_revision 11-Apr-1997 #text_change 03-Dec-1999
C;Accession: A34924; S03292
R;Bourcade, C.; Meissner, D.R.; Atkinson, J.P.; Zeldes, W.; Atkinson, J.P.
J. Biol. Chem. 265, 974-980, 1990
A;Title: Duplication and divergence of the amino terminal coding region of the complement
A;Reference number: A34924; MUID:90110163
A;Accession: A34924
A;Molecule type: DNA
A;Residues: 1-479 <BOU>
A;Cross-references: GB:J05195
R;Bourcade, D.; Meissner, D.R.; Atkinson, J.P.; Holers, V.M.
J. Exp. Med. 168, 1255-1270, 1988
A;Title: Identification of an alternative polyadenylation site in the human C3b/C4b rece
type 1
A;Reference number: S03291; MUID:89010527
A;Accession: S03292
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 'IPPPEAPR',452-482 <BO2>
A;Cross-references: EMBL:X14360
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1988
C;Superfamily: C4b binding protein alpha chain; complement factor H repeat homology
F;43-90/Domain: complement factor H repeat homology ·FH01·
F;104-161/Domain: complement factor H repeat homology ·FH02·
F;166-232/Domain: complement factor H repeat homology ·FH03·
F;238-294/Domain: complement factor H repeat homology ·FH04·
F;297-353/Domain: complement factor H repeat homology ·FH05·
F;358-416/Domain: complement factor H repeat homology ·FH06·
F;421-478/Domain: complement factor H repeat homology ·FH07·

Query Match 21.9%; Score 164.5; DB 2; Length 482;
Best Local Similarity 30.4%; Pred. No. 4.5e-08;
Matches 42; Conservative 19; Mismatches 58; Indels 19; Gaps 5;

Qy 2 ISCDPPPEVKNA RKRYYSLPIVPGTVLRYTCS --- PSYRLIGEKAIPCISEN 50
Db 164 IDCGLPPIIANGDFTSTSREYFHY GSVVTYRCNIGSRGKKVFELVGEPSIYCISKD 219

Qy 51 QVHAIWDKAPPICESVNKPISCSDPIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSK 110
Db 220 DQVGIWSSPAPQCILPNK CPPPNVENGILVSDNRSLFSLNEVVEFRCQPGFIMKGPS 276

Qy 111 IVWCQANEMWGPTALPVC 128
Db 277 HVKCQALNKWEP ELPSC 293

RESULT 28
S65561
factor H - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S65561
R;Soames, C.J.; Day, A.J.; Sim, R.B.
Biochem. J. 315, 523-541, 1996
A;Title: Prediction from sequence comparisons of residues of factor H involved in the
A;Reference number: S65561; MUID:96202005
A;Accession: S65561
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-669 <SOA>
A;Cross-references: GB:X98697; NID:g1419424; PIDN:CAA67257.1; PID:g1419424
C;Superfamily: complement factor H; complement factor H repeat homology
F;55-114/Domain: complement factor H repeat homology ·FHR1·
F;296-349/Domain: complement factor H repeat homology ·FHR2·
F;355-412/Domain: complement factor H repeat homology ·FHR3·
F;416-471/Domain: complement factor H repeat homology ·FHR4·
F;476-530/Domain: complement factor H repeat homology ·FHR·
F;538-592/Domain: complement factor H repeat homology ·FHR5·
F;599-651/Domain: complement factor H repeat homology ·FHR6·

Query Match 20.9%; Score 152.5; DB 2; Length 669;
Best Local Similarity 31.3%; Pred. No. 2.9e-07;
Matches 42; Conservative 14; Mismatches 51; Indels 27; Gaps 6;

Qy 3 SCDPPPPEVKNA-RKRYYSLPIVPGTVLRYTCSPSYRLIGEKAIPCISENQVHAIWDKAPP 61
Db 547 SCALPPELPNGKPKEKETHKFEYAHNEVVEYACNPKFTMKGSHKIQCVDGEST ALP 590

Qy 62 ICESVNKPISCSD-- --PIVPGGPMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWC 114
Db 591 VCIEEERICEISDLDHGDVKPSVP -- PTHRDGSVEFSCKEAFTMIGPKEIIC 640

Qy 115 QANEMWGPTALPVC 128
Db 641 ISGEPW TQPPQC 654

RESULT 29
A44877
C4b binding protein beta chain precursor - human
N;Alternate names: protein S-binding chain, C4b
C;Species: Homo sapiens (man)
C;Date: 20-Jul-1993 #sequence_revision 28-Jul-1993 #text_change 21-Jul-2000
C;Accession: A47107; A44877; A31117
R;Hillarp, A.; Pardo-Manuel, F.; Ruiz, R.R.; Rodriguez de Cordoba, S.; Dahlbaeck, B
J. Biol. Chem. 268, 15017-15023, 1993
A;Title: The human C4b-binding protein beta-chain gene.
A;Reference number: A47107; MUID:93315479
A;Accession: A47107
A;Molecule type: DNA
A;Residues: 1-252 <HI2>
A;Cross-references: GB:L11244; NID:g179682; PID:g387660; GB:L11245; NID:g179683; PID:
R;Hillarp, A.; Dahlbaeck, B.
Proc. Natl. Acad. Sci. U.S.A. 87, 1183-1187, 1990
A;Title: Cloning of cDNA coding for the beta chain of human complement component C4b
A;Reference number: A44877; MUID:90148962
A;Accession: A44877
A;Molecule type: mRNA
A;Residues: 1-252 <HIL>
A;Cross-references: GB:M29964; NID:g179678; PIDN:AAB59520.1; PID:g179679
R;Hillarp, A.; Dahlbaeck, B.
J. Biol. Chem. 263, 12759-12764, 1988
A;Title: Novel subunit in C4b binding protein required for protein S binding.
A;Reference number: A31117; MUID:88315081
A;Accession: A31117
A;Molecule type: protein
A;Residues: 18-20,'PQ',23,26-27,'L',29,'D',31,'VD',34-36, 'F' <HIL>
C;Genetics:
A;Gene: GDB:C4BB
A;Cross-references: GDB:125208; OMIM:120834

```
A;Map position: 1q42 1q42
C;Superfamily: complement factor H repeat homology
C;Keywords: extracellular protein; plasma
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-252/Product: C4b binding protein beta chain #status experimental <MAT>
F;24-76/Domain: complement factor H repeat homology <FH01>
F;81-134/Domain: complement factor H repeat homology <FH02>
F;149-191/Domain: complement factor H repeat homology <FH03>

  Query Match             20.8%;  Score 156.5;  DB 2;  Length 252;
  Best Local Similarity   28.9%;  Pred. No. 1.3e-07;
  Matches   47;  Conservative   19;  Mismatches   57;  Indels   15;  Gaps    6;

Qy       4 CDPPPEVKNAKKPYYSLPIVPGTVL RYTCSPSYRLIGEKAIFCISENQVHATWDKAPPI 62
           |   | | |:    :    | | :|  | |   | |:|:| :|| :  :    ||
Db      23 CPELPPVDNS-  IFVAKPVEGQIIGIYVCIKGYHLVGKKTLFCNASKE ---WDNTFTE 75

Qy      63 CPSVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVIFTCKANFTMKGSKTVWCQANEMWGP 122
           |    :  | ||:: :|  :  | |    | :|| |  :: :|||   |  :  | |
Db      76 C-   RLGHCPDPVLVNGEFS- SSGPVNVSDKITFMCNDHYILKGSNRSQCLEDHTWAP 129

Qy     123 IALPVCES 130
             |:|:|
Db     130 PFPICKS 136


RESULT 30
BBBU
Complement factor B precursor [validated] - human
N;Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-
N;Contains: alternative complement pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment
C;Species: Homo sapiens (man)
C;Date: 19-Feb-1984 #sequence_revision 05-Aug-1994 #text_change 08-Dec-2000
C;Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628; I54
R;Mejia, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.
submitted to the EMBL Data Library, March 1993
A;Reference number: S34075
A;Accession: S34075
A;Molecule type: mRNA
A;Residues: 1-764 <MEJ>
A;Cross-references: EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g297569
R;Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.
Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982
A;Title: Isolation of cDNA clones for the human complement protein factor B, a class III
A;Reference number: A44622; MUID:83039428
A;Accession: A44622
A;Molecule type: mRNA
A;Residues: 467-546;550-595;752-764 <WOO>
A;Cross-references: GB:J00185; GB:J00186
A;Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation
R;Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.
J. Biol. Chem. 259, 3407-3412, 1984
A;Title: Complete primary structure for the zymogen of human complement factor B.
A;Reference number: A20751; MUID:84161997
A;Accession: A00934
A;Molecule type: protein; mRNA
A;Residues: 26-764 <MOL>
A;Cross-references: GB:K01566
A;Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-
A;Note: 736 Ser was also found
A;Note: glycosylation sites were determined
R;Christie, D.L.; Gagnon, J.
Biochem. J. 209, 61-70, 1983
A;Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th
A;Reference number: A19188; MUID:83204092
A;Contents: the final paper in a series documenting the sequence, glycosylation site, an
A;Accession: A19188
A;Molecule type: protein
A;Residues: 260-296,'T',298-764 <CHR>
R;Campbell, R.D.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983
A;Title: Molecular cloning and characterization of the gene coding for human complement
A;Reference number: A19947; MUID:84273641
A;Accession: A19947
A;Molecule type: DNA
A;Residues: 346-764 <CAM>
A;Cross-references: GB:J00125
A;Accession: B19947
A;Molecule type: mRNA
A;Residues: 339-509 <CA1>
A;Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536
R;Wu, L.; Morley, B.J.; Campbell, R.D.
Cell 48, 331-342, 1987
A;Title: Cell-specific expression of the human complement protein factor B gene: evid
A;Reference number: A25971; MUID:87102880
A;Accession: B25971
A;Molecule type: DNA
A;Residues: 1-99 <WU1>
A;Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534
R;Niemann, M.A.; Bhown, A.S.; Miller, E.J.
Biochem. J. 274, 473-480, 1991
A;Title: The principal site of glycation of human complement Factor B.
A;Reference number: S14339; MUID:91174758
A;Accession: S14339
A;Molecule type: protein
A;Residues: 270-329 <NIE>
A;Note: binding site for carbohydrate to lysine under artificial conditions
R;Morley, B.J.; Campbell, R.D.
EMBO J. 3, 153-157, 1984
A;Title: Internal homologies of the Ba fragment from human complement component facto
A;Reference number: A44628; MUID:84158524
A;Accession: A44628
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 16-225,'P',227-259 <MOR>
R;Schwaeble, W.; Luttig, B.; Sokolowski, T.; Staller, C.; Weiss, E.H.; Meyer zum Bus
Immunobiology 188, 221-232, 1993
A;Title: Human complement factor B: functional properties of a recombinant zymogen of
A;Reference number: I54409; MUID:94041399
A;Accession: I54409
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-764 <RIE>
A;Cross-references: GB:S67310; NID:g452937; PIDN:AAB13989.1; PID:g4261689
R;Horiuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.
Mol. Immunol. 30, 1587-1592, 1993
A;Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic c
A;Reference number: I57824; MUID:94067177
A;Accession: I57824
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-31,'J',33-764 <RE2>
A;Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922
C;Comment: 292 Cys has a free sulfhydryl.
C;Genetics:
A;Gene: GDB:BF
A;Cross-references: GDB:119726; OMIM:138470
A;Map position: 6p21.3-6p21.3
A;Introns: 21/3; 94/1; 346/1; 390/1; 414/1; 478/1; 532/3; 542/1; 599/2; 624/1; 652/3;
A;Note: the list of introns may be incomplete
A;Note: gene is located in the major histocompatibility complex, class III region
C;Complex: complement factor B initially forms an inactive complex with complement fa
ment factor C3b forming active C3/C5 convertase; Ba is released
C;Function:
A;Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alph
al
A;Pathway: complement alternate pathway
C;Superfamily: complement C2; complement factor H repeat homology; trypsin homology;
C;Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hyd
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-764/Product: complement factor B #status experimental <MAT>
F;26-259/Product: complement factor Ba fragment #status experimental <BAF>
F;37-98/Domain: complement factor H repeat homology <FH1>
F;103-158/Domain: complement factor H repeat homology <FH2>
F;165-218/Domain: complement factor H repeat homology <FH3>
F;260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>
```

```
F;268-438/Domain: von Willebrand factor type A repeat homology <VFA>
F;482-752/Domain: trypsin homology #status atypical <TRY>
F;97-76,112-98,103-145,131-158,165-205,191-218,478-596,511-527,599-615,656-682,695-725/Di
F;122-141/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;459-460/Cleavage site: Arg-Lys (complement factor D) #status experimental
F;267-276/Active site: His, Asp, Ser #status experimental

  Query Match             20.4%;  Score 154.5;  DB 1;  Length 764;
  Best Local Similarity   41.1%;  Pred. No. 8.2e-07;
  Matches   42;  Conservative   12;  Mismatches   59;  Indels   17;  Gaps    6;

Qy        2 ISCDPPPEVKNA     RKFYYKIFIVFGIVLRYGCSPSYRLIGEKAIFCISENQVHAIWD 57
            | :| |   :|       | |||::         : : |   | | |       |    ||:  |
Db      201 IRCPRPPHDFENGEAYWFKSFYYNV     SDEISFHCYDGYILPCSANPRS  QVNCPWS 252

Qy       58 KAPPICESVNKLISCSGFIVFGKEMNKGSKAFERHGDSVTFTCKANFTMKGSKIVWCQAN 117
             ||:  |   |||:|:| |    ||:  :|  ||||: |    |:::||:   ||
Db      253 GQFTAICD  NGAGYCSNRGIPIGTRKVGSQ  YRLEDSVTYHCSRGLTLRGSQRRTCQEG 208

Qy      118 EMWGPTALPVCESDF 132
              |  |  | |:  |
Db      207 GSWSGTE PSCQDSF 222


RESULT 4
I55975
X/Y protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999
C;Accession: I55975
R;Aggeler-Shaw, M.; Cole, J.L.; Kliickstein, L.B.; Wong, W.W.; Fearon, D.T.; Lalley, P.A
J. Immunol. 148, 4488-4494, 1992
A;Title: Expansion of the complement receptor gene family: identification in the mouse c
A;Reference number: I55975; MUID:92196375
A;Accession: I55975
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-330 <RES>
A;Cross-references: GB:M61179; NID:g202427; PIDN:AAA40574.1; PID:g202428
C;Superfamily: complement factor H repeat homology
F;9-62/Domain: complement factor H repeat homology <FH1>
F;188-244/Domain: complement factor H repeat homology <FH2>
F;249-306/Domain: complement factor H repeat homology <FH3>

  Query Match             19.9%;  Score 150;  DB 2;  Length 330;
  Best Local Similarity   28.2%;  Pred. No. 7.3e-07;
  Matches   37;  Conservative   22;  Mismatches   50;  Indels   22;  Gaps    6;

Qy        4 CPPPEVKNA      RKFYYKIFIVFGIVLRYGCSPSYRLIGEKAIFCISENQVHAIWDK 58
            | :|| | ||       :  : :|       :: : | | :  : |  :: | | |:    |:
Db       46 CIPPPYVENAVMLSENRSLFSI     RDIVEFRCHPGFIMKGASSVHCQSLNK----WEP 87

Qy       59 APPICESVNKLISCSGFIVFGKEM NKGSKAFERHGDSVTFTCKANFTMKGSKI --  V 112
              | |    | : | |    ||     | |  : :|::||  |:  :|:::||        :
Db       88 ELPSC    FKGVTCRLPQEMVEQKGIGMKKEYYYGENVILECEDGYILEGSSQASAIYL 144

Qy      113 WCQANEMWGPT 123
              |    | ||  |
Db      145 WVSGEAWGSI 155


RESULT 5
NBHUH
complement factor H precursor, long splice form [validated] - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-2000
C;Accession: S00254; A00238; A54726; A61565; A26505; I72654; S66298
R;Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A;Title: The complete amino acid sequence of human complement factor H.
A;Reference number: S00254; MUID:88134059
A;Accession: S00254
A;Molecule type: mRNA
A;Residues: 1-1231 <RIP>
A;Cross-references: EMBL:Y00716; NID:g41964; PIDN:CAA68704.1; PID:g41965
A;Note: 402 Tyr was also found
A;Note: parts of this sequence, including the amino and carboxy- ends of the mature p
R;Estaller, C.; Schwaeble, W.; Dierich, M.; Weiss, E.H
Eur. J. Immunol. 21, 799-802, 1991
A;Title: Human complement factor H: two factor H proteins are derived from alternativ
A;Reference number: A60238; MUID:91184292
A;Accession: A60238
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-56;1177-1291 <EST>
A;Note: only portions of this 4.3 kilobase mRNA were sequenced
R;Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biosci. Rep. 7, 201-207, 1987
A;Title: Sequence analysis of a cDNA clone encoding the C terminal end of human compl
A;Reference number: A54726; MUID:88025472
A;Accession: A54726
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'DFRN',579-1231 <DAY>
A;Cross-references: GB:M17517; NID:g180497; PIDN:AAA52015.1; PID:g180498
A;Note: parts of this sequence were determined by protein sequencing
R;Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.I.; Campbell, R.D.; Sim, R.B.
Biosci. Rep. 6, 65-72, 1986
A;Title: Partial characterization of human complement factor H by protein and cDNA se
A;Reference number: A61565; MUID:86188123
A;Accession: A61565
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'METCPNHLNAKI',1050-1052,'T',1054-1102 <RI2>
R;Sim, R.B.; DiScipio, R.G.
Biochem. J. 205, 285-293, 1982
A;Title: Purification and structural studies on the complement system control protein
A;Reference number: A26505; MUID:83048213
A;Accession: A26505
A;Molecule type: protein
A;Residues: 19-29,'Q',22-29,'V',31-34,'Q',35 <SIM>
R;Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Feates, J.; Driscoll, P
Biochemistry 31, 3626-3634, 1992
A;Title: Solution structure of the fifth repeat of factor H: a second example of the
A;Reference number: A44551; MUID:92242649
A;Contents: annotation; NMR structure determination; residues 264-292
R;Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A;Title: Three-dimensional structure of a complement control protein module in soluti
A;Reference number: A49224; MUID:91278097
A;Contents: annotation; NMR structure determination; residues 927-985
R;Estaller, C.; Kolstinen, V.; Schwaeble, W.; Dierich, M.P.; Weiss, E.H
J. Immunol. 146, 3190-3196, 1991
A;Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no

A;Reference number: I56100; MUID:91201892
A;Accession: I72654
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1047-1231 <RES>
A;Cross-references: GB:M65294; NID:g183766; PIDN:AAA35644.1; PID:g183767
R;Carton, J.A.; Bates, R.C.; Smith, A.L.; Ictor, L.; Arellano, A.; Gordon, D.L.; Born
Biochim. Biophys. Acta 1289, 305-311, 1996
A;Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A;Reference number: S66298; MUID:96205365
A;Accession: S66298
A;Status: preliminary
A;Molecule type: protein
A;Residues: 411-419;574-578,580-582 <CAR>
C;Comment: Factor H has also been found bound to cell membranes in an unknown manner.
C;Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liv
C;Genetics: <HFL>
A;Gene: GDB:HFL; HF
A;Cross-references: GDB:120041; OMIM:134370
A;Map position: 1q32-1q32
C;Genetics: <HF2>
```

```
A;Gene: GDB:HF2; HF
A;Cross references: GDB:129095
A;Map position: 1q32-1q42
A;Note: the correspondence between the two loci and the sequences indicated is unclear;
C;Function:
A;Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa
he alternative complement pathway
A;Pathway: complement alternate pathway
C;Superfamily: complement factor H; complement factor H repeat homology
C;Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1229/Product: complement factor H #status experimental <MAT>
F;19-449/Product: complement factor H, short splice form #status experimental <MAT>
F;21-80/Domain: complement factor H repeat homology <FH01>
F;85-141/Domain: complement factor H repeat homology <FH02>
F;146-205/Domain: complement factor H repeat homology <FH03>
F;210-262/Domain: complement factor H repeat homology <FH04>
F;246-248/Region: cell attachment (R-G-D) motif
F;267-320/Domain: complement factor H repeat homology <FH05>
F;325-385/Domain: complement factor H repeat homology <FH06>
F;389-442/Domain: complement factor H repeat homology <FH07>
F;448-505/Domain: complement factor H repeat homology <FH08>
F;509-564/Domain: complement factor H repeat homology <FH09>
F;569-623/Domain: complement factor H repeat homology <FH10>
F;630-684/Domain: complement factor H repeat homology <FH11>
F;691-744/Domain: complement factor H repeat homology <FH12>
F;753-803/Domain: complement factor H repeat homology <FH13>
F;811-864/Domain: complement factor H repeat homology <FH14>
F;870-926/Domain: complement factor H repeat homology <FH15>
F;931-984/Domain: complement factor H repeat homology <FH16>
F;989-1043/Domain: complement factor H repeat homology <FH17>
F;1048-1102/Domain: complement factor H repeat homology <FH18>
F;1109-1164/Domain: complement factor H repeat homology <FH19>
F;1167-1228/Domain: complement factor H repeat homology <FH20>
F;21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357
1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077-
F;217/Binding site: carbohydrate (Asn) (covalent) #status absent
F;529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             19.8%;  Score 149;  DB 1;  Length 1231;
  Best Local Similarity   32.6%;  Pred. No. 3.6e-06;
  Matches   43;  Conservative   18;  Mismatches   49;  Indels   22;  Gaps    9;

Qy        4 SCDPPPE---VKNARKPVYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
            || ||||    ||   |  |         |: | |:| : : |   | |     |   |
Db      629 SCGPPPELLNGNVKEKTKEEYG---HSEVVEYYCNPRFLMKGPNKIQC-----VDGEWT 679

Qy       58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQA 116
             |:|   : :   :| | | :  |:    |  |: :|||| | |  :||| | ::: |
Db      680 TL PVC--IVEESTCGDIPELEHGWAQL-SSPPYYYGDSVEFNCSESFTMIGHRSITC-I 734

Qy      117 NEMWGPTALPVC 128
            : :|  | || |
Db      735 HGVW- TQLPQC 744


RESULT 33
I46709
endothelial leukocyte adhesion molecule 1 - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 19-May-2000
C;Accession: I46709; I46708
R;Lariqan, J.D.; Tsanq, T.C.; Rumberqer, J.M.; Burns, D.K.
DNA Cell Biol. 11, 149-162, 1992
A;Title: Characterization of cDNA and qenomic sequences encoding rabbit ELAM-1: Conserva
A;Reference number: I46708; MUID:92189729
A;Accession: I46709
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-551 <LAR1>
A;Cross-references: GB:M91005; NID:q165006; PIDN:AAA31244.1; PID:q165007
A;Accession: I46708
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-307,'T',309-327,'T',329-490,'V',492-551 <LAR2>
A;Cross-references: GB:M91004; NID:q165004; PIDN:AAA31243.1; PID:q165005
C;Genetics:
A;Gene: ELAM1
A;Map position: 1q22-q25
A;Introns: 13/1; 143/1; 179/1; 241/1; 304/1; 366/1; 429/1; 488/1; 525/1; 549/2
C;Superfamily: unassigned EGF-related proteins; C type lectin homology; complement fa
C;Keywords: duplication; glycoprotein; tandem repeat
F;14-140/Domain: C type lectin homology <LEC>
F;182-239/Domain: complement factor H repeat homology <FH1>
F;244-301/Domain: complement factor H repeat homology <FH2>
F;306-364/Domain: complement factor H repeat homology <FH3>
F;369-427/Domain: complement factor H repeat homology <FH4>
F;432-486/Domain: complement factor H repeat homology <FH5>
F;32,45,201,314,341,466/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             19.7%;  Score 148.5;  DB 2;  Length 551;
  Best Local Similarity   25.9%;  Pred. No. 1.7e-06;
  Matches   36;  Conservative   20;  Mismatches   48;  Indels   35;  Gaps    6;

Qy        1 EISCDPPEVKNARKPVYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 60
            :: | |          |:        |     |   : |:| ::: | |       :||
Db      254 DVKCS----SQGSAPW-------NTTCTFDC-EGFTLIGARSLQCTSS---GSWDNEK 298

Qy       61 PICESV---------NKTISCSDPIVPGGFMNKGSKAPFRHGDSVIFTCKANFTMKGSKT 111
            | |::|         | ::|||:           |:  |    |  |||: || ::|
Db      299 PTCKAVSCDAIHHPQNGSVSCSN------ ---SSEGKFAFRSSCNFTCEENFLLRGPAQ 348

Qy      112 VWCQANEMWGPTALPVCES 130
            | | |   :   | ||||:
Db      349 VECTAQGQWIQQA-PVCEA 366


RESULT 34
A53274
complement factor B subunit Ba - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 02-May-1994 #sequence_revision 18-Nov-1994 #text_change 19-May-2000
C;Accession: A53274
R;Peelman, L.J.; Van de Weghe, A.R.; Coppieters, W.R.; Van Zeveren, A.J.; Bouquet, Y.
Immunogenetics 34, 192-195, 1991
A;Title: Cloning and sequencing of the porcine complement factor B.
A;Reference number: A53274; MUID:91372866
A;Accession: A53274
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-152 <PEE>
A;Note: sequence extracted from NCBI backbone (NCBIN:56029, NCBIP:56042)
C;Superfamily: complement C2; complement factor H repeat homology; trypsin homology;
F;4-59/Domain: complement factor H repeat homology <FH1>
F;66-119/Domain: complement factor H repeat homology <FH2>

  Query Match             19.2%;  Score 144.5;  DB 2;  Length 152;
  Best Local Similarity   30.4%;  Pred. No. 1.1e-06;
  Matches   41;  Conservative   17;  Mismatches   60;  Indels   17;  Gaps    6;

Qy        2 ISCDPPPEVKNA----RKPVYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
            | |  |   :|     | ||||:       : : |   | | |     |    |:   |:
Db        2 IRCPRPHDFENGEYWPRAPYYNL----SLGISFHCYDCYTLRGSANRTC---QVTGRWD 53

Qy       58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPIRHGDSVIFTCKANFTMKGSKTVWCQAN 117
              ||:  :    | :| :| |    |::  :|  |||| :|    |::|:   ||
Db       54 GQTAICD-DGAGYCPNPGIPIGTRKVGSQ--YRLEDSVIYYCTRGLILRGSQRRTCQEG 109

Qy      118 EMWGPTALPVCESDF 132
            |  |  | |:  |
Db      110 GSWSGTE PSCQDSF 123


RESULT 35
A42755
```

```
P Selectin precursor - mouse
N Alternate names: CD62; granule membrane protein 140; PADGEM
C Species: Mus musculus (house mouse)
C Date: 04-Mar-1994 #sequence_revision 18-Nov-1994 #text_change 29-May-2003
C Accession: A42755; A44899
R Weller, A.; Isenmann, S.; Vestweber, D.
J. Biol. Chem. 267, 15176-15183, 1992
A Title: Cloning of the mouse endothelial selectins. Expression of both E- and P-selecti
A Reference number: A42755; MUID:92340571
A Accession: A42755
A Status: nucleic acid sequence not shown
A Molecule type: mRNA
A Residues: 1-708 <WEL>
A Cross-references: GB:M87801; NID:g200552; PIDN:AAA40008.1; PID:g200553
A Experimental source: endothelial cells
A Note: sequence extracted from NCBI backbone (NCBIP:109467)
R Sanders, W.E.; Wilson, R.W.; Ballantyne, C.M.; Beaudet, A.L.
Blood 80, 795-800, 1992
A Title: Molecular cloning and analysis of in vivo expression of murine P-selectin
A Reference number: A44899; MUID:92345617
A Accession: A44899
A Status: nucleic acid sequence not shown
A Molecule type: mRNA
A Residues: 1-723,<E7>,725-788 <SAN>
A Cross-references: GB:M73322; NID:g193565; PIDN:AAA37712.1; PID:g193566
A Note: sequence extracted from NCBI backbone (NCBIP:109900)
C Superfamily: unassigned EGF-related proteins; complement factor H repeat homology; EGF
C Keywords: cell adhesion; glycoprotein; phosphohistidine; phosphoprotein; transmembrane
F 1-41/Domain: signal sequence #status predicted <SIG>
F 42-708/Product: P-selectin #status predicted <MAT>
F 156-194/Domain: EGF homology <EGF>
F 200-257/Domain: complement factor H repeat homology <FH01>
F 262-319/Domain: complement factor H repeat homology <FH02>
F 324-381/Domain: complement factor H repeat homology <FH03>
F 386-443/Domain: complement factor H repeat homology <FH04>
F 448-505/Domain: complement factor H repeat homology <FH05>
F 511-567/Domain: complement factor H repeat homology <FH06>
F 573-630/Domain: complement factor H repeat homology <FH07>
F 634-690/Domain: complement factor H repeat homology <FH08>
F 711-733/Domain: transmembrane #status predicted <TMN>
F 744-768/Domain: intracellular #status predicted <INT>
F 45,54,107,212,347,398,456,462,603,654,661,679/Binding site: carbohydrate (Asn) (covalc

  Query Match              19.1%;  Score 144.5;  DB 2;  Length 768;
  Best Local Similarity    29.5%;  Pred. No. 7.4e-06;
  Matches   83;  Conservative   15;  Mismatches   49;  Indels   15;  Gaps    5;

Qy      29 RYTCSPSYRLIGEKAIFCISENQVHAIWDKAPPICESVNKPLSCSDPEIVD--SGEMNEGS 86
           |  | | ||  |   : |     |     | :  | ||:     |:|  | :|  |     |
Db     289 KFEQGGYRARGSNFIHCTGSQ-     WSEPLPFCEA     IACPPPEIPNCHGSMPCVPS 340

Qy      87 KAPTFFRERVLI--KANFIMKDIPEIVWPSANLMWGPLALFVGL-    DFIL 134
             |  :  | || |   | :||:  : |  :  |   | | ||:     :|||
Db     341 IGIFGYNSSCIFLCAEGFVLKGNDAIQCADSGQWTAPA PFCEALQCPEFPV 391


RESULT 36
WMBEIE
secretory complement control protein homolog precursor - saimiriine herpesvirus 1 (strai
N Alternate names: sCCPH protein
C Species: saimiriine herpesvirus 1
C Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 01-Dec-2000
C Accession: A42534; B36806; S24566
R Albrecht, J.C.; Fleckenstein, B.
J. Virol. 66, 3937-3940, 1992
A Title: New member of the multigene family of complement control proteins in herpesviru
A Reference number: A42534; MUID:92260674
A Accession: A42534
A Molecule type: DNA
A Residues: 1-302 <ALB>
A Cross-references: EMBL:X60283; NID:g60315; PIDN:CAA42822.1; PID:g60316
C Genetics:
A Introns: 288/1
C Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C Keywords: alternative splicing; duplication; extracellular protein; glycoprotein
F 1-20/Domain: signal sequence #status predicted <SIG>
F 21-302/Product: secretory complement control protein homolog #status predicted <SCC
F 23-79/Domain: complement factor H repeat homology <FH1>
F 84-142/Domain: complement factor H repeat homology <FH2>
F 147-205/Domain: complement factor H repeat homology <FH3>
F 210-264/Domain: complement factor H repeat homology <FH4>
F 36,49,46,72,155/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              18.9%;  Score 142.5;  DB 1;  Length 302;
  Best Local Similarity    27.2%;  Pred. No. 3.5e-06;
  Matches   47;  Conservative   18;  Mismatches   64;  Indels   17;  Gaps    5;

Qy       4 CDPPPEVKNAP-    KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDK 58
           |  | ::  | |        ||       |:|: |||:  | |||     |: :     |
Db      84 CSTPQDLLNGRYIVIGNLYY      GSVITYICNSGYSLIGSTISACLLKRQGRVDWTP 137

Qy      59 APFCESVNKTLSCSDPIVPGGEMNKGSKAPFFRERVLIKANFIMKDIPEIVWPSANLMWGPLA 116
           |||:  : |   | |   |     :   |  : : |:|::|     :  | :  |
Db     138 RPPICD-IKK---CKPPPQIANGTHTNVKDFYTYLDTVIYSCNDETKLTLIGFSSKLCSE 193

Qy     117 NEMWGPTALPVCESDF 132
             | |     ||  |
Db     194 TGSWVPNGEIKCEFIF 209


RESULT 37
WMBE2E
membrane-bound complement control protein homolog precursor - saimiriine herpesvirus
N Alternate names: mCCPH protein
C Species: saimiriine herpesvirus 1
C Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C Accession: B42534, D36806; S24567
R Albrecht, J.C.; Fleckenstein, B.
J. Virol. 66, 3937-3940, 1992
A Title: New member of the multigene family of complement control proteins in herpesv
A Reference number: A42534; MUID:92260674
A Accession: B42534
A Molecule type: DNA
A Residues: 1-460 <ALB>
A Cross-references: EMBL:X60283; NID:g60315; PIDN:CAA42823.1; PID:g60317
C Superfamily: herpesvirus complement control protein; complement factor H repeat hom
C Keywords: duplication; glycoprotein; transmembrane protein
F 1-20/Domain: signal sequence #status predicted <SIG>
F 21-302/Product: membrane-bound complement control protein homolog #status predicted
F 23-79/Domain: complement factor H repeat homology <FH1>
F 84-142/Domain: complement factor H repeat homology <FH2>
F 147-205/Domain: complement factor H repeat homology <FH3>
F 210-264/Domain: complement factor H repeat homology <FH4>
F 329-345/Domain: transmembrane #status predicted <TMN>
F 36,49,46,72,155,294/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              18.9%;  Score 142.5;  DB 1;  Length 460;
  Best Local Similarity    27.2%;  Pred. No. 4.2e-06;
  Matches   47;  Conservative   18;  Mismatches   64;  Indels    7;  Gaps    5;

Qy       4 CDPPPEVKNAP- ---KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDK 58
           |  | ::  | |        ||       |:|: |||:  | |||     |: :     |
Db      84 CSTPQDLLNGRYIVIGNLYY      GSVITYICNSGYSLIGSTISACLLKRQGRVDWTP 137

Qy      59 APFCESVNKTLSCSDPIVPGGEMNKGSKAPFFRERVLIKANFIMKDIPEIVWPSANLMWGPLA 116
           |||:  : |   | |   |     :   |  : : |:|::|     :  | :  |
Db     138 RPPICD-IKK  -CKPPPQIANGTHTNVKDFYTYLDTVIYSCNDETKLTLIGFSSKLCSE 193

Qy     117 NEMWGPTALPVCESDF 132
             | |     ||  |
Db     194 TGSWVPNGEIKCEFIF 209


RESULT 38
H26359
decay-accelerating factor, GPI-anchored splice form precursor - human
```

```
N;Alternate names: CD55; DAF splice form 2; decay-accelerating factor membrane-bound for
C;Species: Homo sapiens (man)
C;Date: 05-Oct 1988 #sequence_revision 16-Aug-1996 #text_change 19-Jan-2001
C;Accession: B26459; A27666; A39101; I52594; I52564
R;Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr , D W.; Nussenzweig, V
Nature 325, 545-549, 1987
A;Title: Cloning of decay accelerating factor suggests novel use of splicing to generate
A;Reference number: A26359; MUID:87115845
A;Accession: B26459
A;Molecule type: mRNA
A;Residues: 1-381 <CAR>
A;Cross-references: GB:M30142; NID:g181464; PIDN:AAA52168.1; PID:g181465
R;Medof, M.E.; Lublin, D.M.; Holers, V.M.; Ayers, D.J.; Getty, R.R.; Leykam, J.F.; Atkin
Proc. Natl. Acad. Sci. U.S.A. 84, 2007-2011, 1987
A;Title: Cloning and characterization of cDNAs encoding the complete sequence of decay-a
A;Reference number: A27666; MUID:87175602
A;Accession: A27666
A;Molecule type: mRNA
A;Residues: 6-79,'T',81-84,'M',86-381 <MED>
A;Cross-references: GB:M15799; NID:g181462; PIDN:AAA52167.1; PID:g181463
R;Moran, P.; Raab, H.; Kohr, W.J.; Caras, I.W.
J. Biol. Chem. 266, 1250-1257, 1991
A;Title: Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavag
A;Reference number: A39101; MUID:91093238
A;Accession: A39101
A;Molecule type: protein
A;Residues: 338-352 <MOR>
R;Lublin, D M ; Mallinson, G ; Poole, J ; Reid, M E ; Thompson, E S ; Ferdman, B.R.; Tel
Blood 84, 1276-1282, 1994
A;Title: Molecular basis of reduced or absent expression of decay-accelerating factor in
A;Reference number: I52594; MUID:94425573
A;Accession: I52594
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 194-198,'L',200-209 <LUB>
A;Cross-references: GB:S72858; NID:g639599; PIDN:AAC60633.1; PID:g639600
A;Experimental source: individual KW, Cromer blood group phenotype Dr(a-)
A;Note: the single nucleotide difference in this allele, which changes Ser-199 to Leu, l
on (see reference I52564), and thus reduced DAF expression
R;Reid, M.E.; Mallinson, G., Sim, R.B., Poole, J., Pausch, V , Merry, A H , Liew, Y.W.,
Blood 78, 3291-3297, 1991
A;Title: Biochemical studies on red blood cells from a patient with the Inab phenotype (
A;Reference number: I52564; MUID:92075980
A;Accession: I52564
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 190-193,'QLCPVE' <RE2>
A;Cross-references: GB:S70688; NID:g240301; PIDN:AAB20576.1; PID:g240302
A;Experimental source: individual KW, Cromer blood group phenotype Dr(a-) (described ing
C;Comment: Cromer blood group system antigens reside on this protein.
C;Comment: For an alternative splice form, see PIR:A26359.
C;Genetics:
A;Gene: GDB:DAF
A;Cross-references: GDB:119088; OMIM:125240
A;Map position: 1q32-1q32
C;Function:
A;Description: protects tissues from damage by regulating complement activation on cell
C;Superfamily: decay-accelerating factor; complement factor H repeat homology
C;Keywords: alternative splicing; blocked carboxyl end; complement inhibitor; glycoprote
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-353/Product: decay-accelerating factor 2 #status predicted <MAT>
F;36-94/Domain: complement factor H repeat homology <FH01>
F;98-158/Domain: complement factor H repeat homology <FH02>
F;163-220/Domain: complement factor H repeat homology <FH03>
F;225-283/Domain: complement factor H repeat homology <FH04>
F;354-381/Domain: carboxyl-terminal propeptide #status predicted <PRO>
F;95/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;353/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form

  Query Match             18.9%;  Score 142;  DB 1;  Length 381;
  Best Local Similarity   29.9%;  Pred. No. 4.9e-06;
  Matches   40;  Conservative   24;  Mismatches   50;  Indels   20;  Gaps    9;

Qy        3 SCDPPPEVKNAKKPYYSLG--IVPGTVLRYTCSPSYRL-GEKAIFC ISENQVHAIWDKA 59
            ||  | |::| :     :|  |: |  : ::    |: |  : || || : |
Db      162 SCPNPGEIRNGQ---IDVPGGILFGATISFSCNTGYKLFGSTSSFCLISGSSVQ  WSDP 216

Qy       60 PPICESVNKTISC-SDPIVPGGFMNKGSKAPFPHGDSVTFTCKANFTMKGSKTVWCQAN  117
             | |    : | | : |  : | : :|     :  | |: |   ||| |  ::: |
Db      217 LPEC----KEIYCPAPPQIDNGII-QGERDHYGYRQSVTYACNKGFTMIGEHSIYCIVNN 271

Qy      118 ---EMWGPTALPVC 128
              |  ||   | |
Db      272 DEGEWSGPP--PEC 283


RESULT 39
A26359
decay accelerating factor, splice form 1 precursor - human
N;Alternate names: decay-accelerating factor 5; decay-accelerating factor secreted fo
C;Species: Homo sapiens (man)
C;Date: 05 Oct 1988 #sequence_revision 05-Oct-1988 #text_change 02-Jun 2000
C;Accession: A26359, A39702; S16187; S23138; A27258
R;Caras, I.W.; Davitz, M.A.; Rhee, L.; Weddell, G.; Martin Jr., D.W.; Nussenzweig, V.
Nature 325, 545-549, 1987
A;Title: Cloning of decay-accelerating factor suggests novel use of splicing to gener
A;Reference number: A26359; MUID:87115845
A;Accession: A26359
A;Molecule type: mRNA
A;Residues: 1-440 <CAR>
A;Cross-references: GB:M30142
R;Ewulonu, U.K.; Ravi, L.; Medof, M.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 4675-4679, 1991
A;Title: Characterization of the decay-accelerating factor gene promoter region
A;Reference number: A39702; MUID:91271256
A;Accession: A39702
A;Molecule type: DNA
A,Residues: 1-79,'T',81-104 <EWU>
A,Cross-references: GB:M64356
A,Note: the authors translated the codon ACT for residue 85 as Met
R;Nakano, Y.; Sugita, Y.; Ishikawa, Y.; Choi, N.H.; Tobe, T.; Tomita, M.
Biochim. Biophys. Acta 1074, 326-330, 1991
A,Title: Isolation of two forms of decay-accelerating factor (DAF) from human urine
A;Reference number: S16187; MUID:91291869
A;Accession: S16187
A;Status: preliminary
A;Molecule type: protein
A;Residues: 35-47 <BIO>
R;Nakano, Y.; Simida, K.; Kikuta, N.; Miura, N.H.; Tobe, T.; Tomita, M
Biochim. Biophys. Acta 1116, 235-240, 1992
A,Title: Complete determination of disulfide bonds localized within the short consens
A;Reference number: S23138; MUID:92305084
A;Accession: S23138
A;Status: preliminary
A;Molecule type: protein
A;Residues: 35-41-65-68;79-81;93-103;128-131;143-145;155-159;162-168;188-192;204-204;
R;Sugita, Y.; Nagoro, T.; Matsuda, T.; Sakamoto, T.; Tomita, M.
J. Biochem. 100, 143-150, 1986
A;Title: Improved method for the isolation and preliminary characterization of human
A;Reference number: A27258; MUID:87008481
A;Accession: A27258
A;Molecule type: protein
A;Residues: 35, 'X',37,'G',39-51,'P',53-55, 'X',57-58,'X',60-63 <SUG>
A;Note: Gly 35 and Leu 38 were also found
C;Comment: For an alternative splice form, see PIR:H26359
C;Genetics:
A,Gene: GDB:DAF
A,Cross-references: GDB:119088, OMIM:125240
A;Map position: 1q32-1q32
C;Superfamily: decay-accelerating factor; complement factor H repeat homology
C,Keywords: alternative splicing; glycoprotein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-440/Product: decay accelerating factor 1 #status predicted <MAT>
F,36-94/Domain: complement factor H repeat homology <FH01>
F;98-158/Domain: complement factor H repeat homology <FH02>
F;163-220/Domain: complement factor H repeat homology <FH03>
```

F;2-57/Domain: complement factor H repeat homology <FH04>
F;95/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.6%; Score 142; DB 2; Length 440;
Best Local Similarity 29.9%; Pred. No. 5.7e-06;
Matches 40; Conservative 24; Mismatches 50; Indels 20; Gaps 9;

Qy 3 SCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFC ISENQVHATWDKA 59
Db 162 SCPNPGETRNGQ IDVPGGILFGATISFSCNTGYKLFGSTSSFCLISGSSVQ WSDP 216

Qy 60 PPICESVNKIISC SDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQAN 117
Db 217 LPEC RETYCPAPPQIDNGII QGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVNN 271

Qy 118 EMWGPTALPVC 128
Db 272 DEGEWSGPP PEC 283

RESULT 40
S36772
E-selectin - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change 19-May-2000
C;Accession: S36772
R;Nguyen, M.; Strubel, N.A.; Bischoff, J.
Nature 365, 267-269, 1993
A;Title: A role for sialyl Lewis X/A glycoconjugates in capillary morphogenesis.
A;Reference number: S36772; MUID:93382537
A;Accession: S36772
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-485 <NGU>
A;Cross-references: GB:L12059; NID:g402913; PIDN:AAA02991.1; PID:g402914
C;Superfamily: unassigned EGF related proteins; C type lectin homology; complement facto
F;1-139/Domain: C type lectin homology <LCH>
F;181-232/Domain: complement factor H repeat homology <FH1>
F;234-290/Domain: complement factor H repeat homology <FH2>
F;334-386/Domain: complement factor H repeat homology <FH3>
F;387-421/Domain: complement factor H repeat homology <FH4>

Query Match 18.9%; Score 142; DB 2; Length 485;
Best Local Similarity 29.8%; Pred. No. 6.4e-06;
Matches 42; Conservative 14; Mismatches 55; Indels 48; Gaps 7;

Qy 2 ISCDPPPE VKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
Db 240 VKCDAISNEDNGVVNCSDNHGNSFW NETCIFEFCQPGYKLEGPQHLQCISS ---CIWD 293

Qy 58 KAPPICESV NKIISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKG 108
Db 294 NKQFTCKAVSCAAISHFQNGIVNCSHSVV GDFAFK - SSCHFTCAEGFTLQG 343

Qy 109 SKTVWCQANEMWGPTALPVCE 129
Db 344 PTQVECTAQGQW TQRVPVCE 363

RESULT 41
G35070
apolipoprotein H related protein 1361 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Jul-1990 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
C;Accession: G35070; A35071; B35071; H35070; I45070
R;Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 4193-4201, 1990
A;Title: Identification and sequence analysis of four complement factor H-related transc
A;Reference number: A35070; MUID:90153969
A;Accession: G35070
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-443 <VIK>
A;Cross references: GB:M29008; NID:g192559; PIDN:AAA37414.1; PID:g409165; GB:J05259
A;Note: translation of the nucleotide sequence is not complete
C;Superfamily: apolipoprotein H; complement factor H repeat homology
C;Keywords: duplication
F;28-88/Domain: complement factor H repeat homology <FH1>
F;92-145/Domain: complement factor H repeat homology <FH2>
F;154-208/Domain: complement factor H repeat homology <FH3>
F;215-269/Domain: complement factor H repeat homology <FH4>
F;273-334/Domain: complement factor H repeat homology <FH5>

Query Match 18.6%; Score 139.5; DB 2; Length 443;
Best Local Similarity 30.2%; Pred. No. 7.2e-06;
Matches 39; Conservative 16; Mismatches 61; Indels 13; Gaps 7;

Qy 4 CDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPPI 62
Db 154 CGPPPPIDNGDITSLSLPVYASESSVFYGCEAYYELKGNKTID ENGKWSE PPT 207

Qy 63 CESVNKIISCS-DPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQANEMWG 121
Db 208 C--IYPTGKCGPPPPIDNGDITSISLFYFPLSSVFYGCQNYYVLKGKKTITCQ NGKWS 264

Qy 122 --PTALPVC 128
Db 265 KPPHCLSAC 273

RESULT 42
T42921
complement control protein homolog ceph - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C;Accession: T42921
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: 222274
A;Accession: T42921
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-360 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95530.1
A;Experimental source: strain 73
C;Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Query Match 18.6%; Score 139.5; DB 2; Length 360;
Best Local Similarity 25.4%; Pred. No. 8.1e-06;
Matches 33; Conservative 21; Mismatches 65; Indels 11; Gaps 4;

Qy 4 CDPPPEVKNARKPYYSLP--IVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
Db 84 CSIPADLLNG---WYIVTGNLYYGSVITYTCNTGYQLIGSPTSSCLSPDGKVNWTPRPP 140

Qy 62 ICESVNKIISCSDPIVPGGFMNKGSKAPFRHGDSVTFTC KANFTMKGSKIVWCQANEM 119
Db 141 ICE ITKCKPPPTIANGIHTNIKEYYIYLDAVIYSCNDGYKLLIGESSLQCSETGR 196

Qy 120 WGPTALPVCE 129
Db 197 WVPDEEIKCE 206

RESULT 43
A35046
E-selectin precursor - human
N;Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM 1); LECAM 2
C;Species: Homo sapiens (man)
C;Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 21-Jul-2000
C;Accession: A38615; A35046; A42606
R;Collins, T.; Williams, A.; Johnston, G.I.; Kim, J.; Eddy, R.; Shows, T.; Gimbrone J
J. Biol. Chem. 266, 2466-2473, 1991
A;Title: Structure and chromosomal location of the gene for endothelial leukocyte adh

A;Reference number: A48615; MUID:91115870
A;Accession: A48615
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-610 <COL>
A;Cross-references: GB:M61893; GB:M58017; NID:g182043; PIDN:AAA52375.1; PID:g182046
R;Hession, C.; Osborn, L.; Goff, D.; Chi Rosso, G.; Vassallo, C.; Pasek, M.; Pittack, C.
Proc. Natl. Acad. Sci. U.S.A. 87, 1673-1677, 1990
A;Title: Endothelial leukocyte adhesion molecule 1: direct expression cloning and functi
A;Reference number: A45046; MUID:90175359
A;Accession: A45046
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-610 <HES>
A;Cross-references: GB:M30640; NID:g182047; PIDN:AAA52377.1; PID:g182048
R;Bevilacqua, M.P.; Stengelin, S.; Gimbrone Jr., M.A.; Seed, B.
Science 243, 1160-1165, 1989
A;Title: Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophil
A;Reference number: A32606; MUID:89162047
A;Accession: A32606
A;Molecule type: mRNA
A;Residues: 1-467,'Y',469-610 <BEV>
A;Cross-references: GB:M24736; NID:g537523; PIDN:AAA52376.1; PID:g537524
C;Genetics:
A;Gene: GDB:SELE; ELAM; ESEL; ELAM1
A;Cross-references: GDB:120612; OMIM:131210
A;Map position: 1q22-1q25
C;Superfamily: unassigned EGF-related proteins; C-type lectin homology; complement factd
C;Keywords: duplication; glycoprotein; tandem repeat; transmembrane protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;12-148/Domain: C-type lectin homology <LCH>
F;22-610/Product: endothelial leukocyte adhesion molecule 1 #status predicted <MAT>
F;148-174/Domain: EGF homology <EGF>
F;180-237/Domain: complement factor H repeat homology <FH01>
F;242-299/Domain: complement factor H repeat homology <FH02>
F;304-362/Domain: complement factor H repeat homology <FH03>
F;367-425/Domain: complement factor H repeat homology <FH04>
F;430-488/Domain: complement factor H repeat homology <FH05>
F;494-547/Domain: complement factor H repeat homology <FH06>
F;557-578/Domain: transmembrane #status predicted <TMM>
F;25,145,160,179,199,203,265,312,332,503,527/Binding site: carbohydrate (Asn) (covalent)

  Query Match 18.5%; Score 139; DB 2; Length 610;
  Best Local Similarity 41.5%; Pred. No. 1.6e-05;
  Matches 44; Conservative 18; Mismatches 36; Indels 20; Gaps 6;

```
Qy      40 YICSPSYRLIGEKAIFCISENQVHAIWDKAPPICESVNKTISCSDPIVPGGFMN--KGSK 87
           :||   : | |   : | :: |    | :  |:||:   | : |:|    |:||    :
Db     343 FTCEEGFMLQGPAQVECTTQGQ    WTQQIPVCEAFQCT-ALSNP--ERGYMNCLPSAS 385

Qy      88 APFRHGDSVTFTCKANFTMKGSKTVWCQANEMWGPTA-----LPVCES 130
           ||:| |   :|:  | :|||| : |      |||       | ||:
Db     386 GSFRYGSSCEFSCEQGFVLKGSKRLQC----  GPTGEWDNEKPTCEA 427
```

RESULT 44
I52657
seizure related protein SEZ-6 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 31 Mar-2000
C;Accession: I52657
R;Shimizu-Nishikawa, K.; Kajiwara, K.; Kimura, M.; Katsuki, M.; Sugaya, E.
Brain Res. Mol. Brain Res. 28, 201-210, 1995
A;Title: Cloning and expression of SEZ-6, a brain specific and seizure related cDNA.
A;Reference number: I52657; MUID:95240392
A;Accession: I52657
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-977 <RES>
A;Cross-references: GB:D29763; NID:g693909; PIDN:BAA06167.1; PID:g693910
C;Superfamily: complement factor H repeat homology
F;454-499/Domain: complement factor H repeat homology <FH01>
F;529-586/Domain: complement factor H repeat homology <FH02>
F;707-762/Domain: complement factor H repeat homology <FH03>
F;768-827/Domain: complement factor H repeat homology <FH04>
F;835-892/Domain: complement factor H repeat homology <FH05>

  Query Match 18.4%; Score 138.5; DB 2; Length 977;
  Best Local Similarity 28.2%; Pred. No. 2.9e-05;
  Matches 35; Conservative 17; Mismatches 57; Indels 15; Gaps 4;

```
Qy       3 SCDPPPEVKNAKKPYYSLPIVPGTVLKYTCSPSYFLIGEKAIFCISENQVHAIWDKAPET 62
           :|   ||: :  |      :| |||: | | | |:::|   : |    |  :  :  |
Db     706 TCPELPETPNGWKNPSQPELVHGTVVTYQCYPGYVVGSSTLMC----QWDLSWSEDLPS 761

Qy      63 CESVNKTISCSDPIVPGGFMNKG----SKAFFRHGDSVTFTCKANFTMKGSKTVWCQANE 118
           |: |    || ||    | :       |       | :| : ,   | : ||    |  :
Db     762 CQRVT   SCHDP    GDVEHSKRLISSPKFFVCATVQYVCDQGFVLIGSAILTCHDRQ 814

Qy     119 MWGP 122
           |
Db     815 AGSP 818
```

RESULT 45
KFHU13
coagulation factor XIII chain b precursor - human
N;Alternate names: fibrin-stabilizing factor; plasma transglutaminase b chain; protei
C;Species: Homo sapiens (man)
C;Date: 19-Apr-1991 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C;Accession: A36397; A23830; S09980
R;Bottenus, R.E.; Ichinose, A.; Davie, E.W.
Biochemistry 29, 11195-11209, 1990
A;Title: Nucleotide sequence of the gene for the b subunit of human factor XIII.
A;Reference number: A36397; MUID:91105054
A;Accession: A36397
A;Molecule type: DNA
A;Residues: 1-661 <BOT>
A;Cross-references: GB:J05294
R;Ichinose, A.; McMullen, B.A.; Fujikawa, K.; Davie, E.W.
Biochemistry 25, 4633-4638, 1986
A;Title: Title: Amino acid sequence of the b subunit of human factor XIII, a protein
A;Reference number: A23830; MUID:87026535
A;Accession: A23830
A;Molecule type: mRNA
A;Residues: 2-661 <ICH>
A;Cross-references: GB:M14057; NID:g182833; PIDN:AAA88042.1; PID:g182839
A;Note: parts of this sequence, including the amino end of the mature protein, were d
R;Grundmann, U.; Nerlich, C.; Rein, T.; Zettlmeiss, G.
Nucleic Acids Res. 18, 2817-2818, 1990
A;Title: Complete cDNA sequence encoding the b subunit of human factor XIII.
A;Reference number: S09980; MUID:90251467
A;Accession: S09980
A;Molecule type: mRNA
A;Residues: 1-661 <GRU>
A;Cross-references: EMBL:X51823
C;Comment: This protein is noncatalytic.
C;Genetics:
A;Gene: GDB:F13B
A;Cross-references: GDB:119893; OMIM:134580
A;Map position: 1q31-1q32.1
A;Introns: 2/1; 69/1; 131/1; 199/2; 244/1; 309/1; 373/1; 438/1; 499/2; 561/2; 615/2
C;Complex: heterotetramer in plasma with coagulation factor XIII chain a; the plasma
C;Superfamily: coagulation factor XIII chain b; complement factor H repeat homology
C;Keywords: blood coagulation; duplication; glycoprotein; plasma
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-661/Product: coagulation factor XIII chain b #status predicted <MAT>
F;25-87/Domain: complement factor H repeat homology <FH01>
F;91-146/Domain: complement factor H repeat homology <FH02>
F;153-208/Domain: complement factor H repeat homology <FH03>
F;213-267/Domain: complement factor H repeat homology <FH04>
F;274-327/Domain: complement factor H repeat homology <FH05>
F;336-389/Domain: complement factor H repeat homology <FH06>
F;396-450/Domain: complement factor H repeat homology <FH07>
F;454-515/Domain: complement factor H repeat homology <FH08>
F;524-578/Domain: complement factor H repeat homology <FH09>

```
F;58, 640/Domain: complement factor H repeat homology -FH10-
F;61, 610/Region: cell attachment (RGD) motif
F;25, 76,56, 87,91 105,118 146,153 197,180 208,214 255,241 267,274 316,302 327,336 378,364
F;162,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              18.2%;  Score 137;  DB 1;  Length 661;
  Best Local Similarity    27.9%;  Pred. No. 2.7e-05;
  Matches   86;  Conservative   19;  Mismatches   62;  Indels   12;  Gaps    5;

Qy        1 EISCDPPPEVKNARKPYYSIDIVPGIVLRYYCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
            :::|  |  ::|     :|     |  :| |  | | |   | |          |  |
Db      333 KVACEEPPEIENGAANLIHSKIYYNGDKVIYACKSGYLLRGSNEITC      NRGKW TLP 386

Qy       61 PICESVNKTISCS DFIVPGGLMNKGSKAFEKHGDSVIFICKANFIMKGSKTVWCQANEM 119
              | |   |:  :|   |:|    :  |  | :  | || : |   : ::|||   |:  :
Db      387 PECVENNE  NCKHPPVVMNGAVAIGTLASYATGSSVEYRCNEYYLLRGSKTSRCEQGKW 444

Qy      120 WGPTALPVC 128
              |   :||
Db      445 SSP   PVC 450


Search Completed  November  6, 2002, 16:07:33
Job Time   17.9589 secs
```

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:04:58 ; Search time 8.5316 Seconds
(without alignments)
612.680 Million cell updates/sec

Title: US-09-834-309-6
Perfect score: 752
Sequence: 1 EISCDPPPEVKNARKPYYSL.....ANEMWGPTALPVCESDFPLE 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	752	100.0	1025	1	CR2_MOUSE	P19070 mus musculu
2	466.5	62.0	1033	1	CR2_HUMAN	P20023 homo sapien
3	225.5	30.0	597	1	C4BP_HUMAN	P04003 homo sapien
4	222	29.5	469	1	C4BP_MOUSE	P08607 mus musculu
5	220.5	29.3	2039	1	CR1_HUMAN	P17927 homo sapien
6	208.5	27.7	558	1	C4BP_RAT	Q63514 rattus norv
7	200	26.6	610	1	C4BP_BOVIN	Q28065 bos taurus
8	193.5	25.7	263	1	VCP_VACCV	P10998 vaccinia vi
9	188.5	25.1	377	1	MCP_HUMAN	P15529 homo sapien
10	176	23.4	1244	1	CFAB_MOUSE	P06909 mus musculu
11	157.5	20.9	685	1	CFAB_BOVIN	Q28085 bos taurus
12	156.5	20.8	252	1	C4BB_HUMAN	P20851 homo sapien
13	153.5	20.4	407	1	DAF2_MOUSE	Q61476 mus musculu
14	153.5	20.4	764	1	CFAB_HUMAN	P00751 homo sapien
15	149	19.8	1241	1	CFAB_HUMAN	P08603 homo sapien
16	148.5	19.7	551	1	LEM2_RABIT	P27113 oryctolagus
17	148	19.7	390	1	DAF1_MOUSE	Q61475 mus musculu
18	144.5	19.2	151	1	CFAB_PIG	Q03710 sus scrofa
19	143.5	19.1	768	1	LEM3_MOUSE	Q01102 mus musculu
20	142.5	18.9	360	1	CCPH_HSVSA	Q01016 herpesvirus
21	142	18.9	481	1	DAF_HUMAN	P08174 homo sapien
22	142	18.9	485	1	LEM2_BOVIN	P98107 bos taurus
23	140	18.6	549	1	LEM2_RAT	P98105 rattus norv
24	139	18.5	610	1	LEM2_HUMAN	P16581 homo sapien
25	137	18.2	661	1	F13B_HUMAN	P05160 homo sapien
26	135	18.0	668	1	F13B_MOUSE	Q07968 mus musculu
27	134	17.8	297	1	APOH_RAT	P26644 rattus norv
28	133.5	17.8	258	1	C4BB_RAT	Q63515 rattus norv
29	131.5	17.5	612	1	LEM2_MOUSE	Q00690 mus musculu
30	130.5	17.4	752	1	CO2_HUMAN	P06681 homo sapien
31	130.5	17.4	761	1	CFAB_MOUSE	P04186 mus musculu
32	130.5	17.4	1019	1	LEC_TACTR	P28175 tachypleus
33	130	17.3	340	1	DAF_PONPY	P49457 pongo pygma
34	129.5	17.2	768	1	LEM3_RAT	P98106 rattus norv
35	129	17.2	611	1	LEM2_CANFA	P43730 canis famil
36	128.5	17.1	769	1	LEM3_SHEEP	P98109 ovis aries
37	127.5	17.0	1019	1	LEC_CARRO	Q25422 carcinoscor
38	127	16.9	345	1	APOH_BOVIN	P17690 bos taurus
39	125.5	16.7	330	1	FHR1_HUMAN	Q03591 homo sapien
40	124.5	16.6	646	1	LEM3_BOVIN	P42201 bos taurus
41	124	16.5	345	1	APOH_MOUSE	Q01339 mus musculu
42	123.5	16.4	484	1	LEM2_PIG	P98110 sus scrofa
43	123	16.4	345	1	APOH_CANFA	P33703 canis famil
44	121	16.1	507	1	DAF_CAVPO	Q60401 cavia porce
45	120	16.0	345	1	APOH_HUMAN	P02749 homo sapien
46	119	15.8	830	1	LEM3_HUMAN	P16109 homo sapien
47	118	15.7	958	1	HIG_DROME	Q09101 drosophila
48	117.5	15.6	202	1	APAR_PIG	Q03472 sus scrofa
49	117	15.6	760	1	CO2_MOUSE	P21180 mus musculu
50	116	15.4	270	1	FHR2_HUMAN	P36980 homo sapien
51	113.5	15.1	843	1	CO7_HUMAN	P10643 homo sapien
52	108.5	14.4	317	1	VB05_VACCV	Q01227 vaccinia vi
53	105.5	14.0	330	1	FHR3_HUMAN	Q02985 homo sapien
54	104.5	13.9	317	1	VB05_VACCO	P24084 vaccinia vi
55	104.5	13.9	317	1	VB05_VACCL	P24083 vaccinia vi
56	104.5	13.9	934	1	CO6_HUMAN	P13671 homo sapien
57	102.5	13.6	317	1	VB05_VACCC	P21115 vaccinia vi
58	100	13.3	464	1	SRPX_RAT	Q64769 rattus norv
59	99.5	13.2	250	1	CFBL_CHICK	P81475 gallus gall
60	99.5	13.2	331	1	FHR4_HUMAN	Q92496 homo sapien
61	97	12.9	464	1	SRPX_HUMAN	P78539 homo sapien
62	95.5	12.7	914	1	PERT_MOUSE	P15419 mus musculu
63	95.5	12.7	914	1	PERT_RAT	P.4650 rattus norv
64	94.5	12.6	372	1	LEM1_RAT	P30836 rattus norv
65	93.5	12.4	933	1	PERT_HUMAN	P07202 homo sapien
66	88.5	11.8	372	1	LEM1_MACMU	Q95198 macaca mula
67	88.5	11.8	372	1	LEM1_PAPHA	Q28768 papio hamad
68	87.5	11.6	926	1	PERT_PIG	P09933 sus scrofa
69	86	11.4	268	1	IL2A_CANFA	Q62802 canis famil
70	85	11.3	198	1	C4BB_BOVIN	Q28066 bos taurus
71	84.5	11.2	370	1	LEM1_BOVIN	P98131 bos taurus
72	84.5	11.2	372	1	LEM1_HUMAN	P14151 homo sapien
73	84.5	11.2	372	1	LEM1_PANTR	Q95237 pan troglod
74	83.5	11.1	372	1	LEM1_PONPY	Q95235 pongo pygma
75	83	11.0	347	1	HPT_MUSSA	Q62558 mus saxicol
76	83	11.0	686	1	MAS2_HUMAN	Q00187 homo sapien
77	80.5	10.7	268	1	IL2A_MOUSE	P01590 mus musculu
78	80.5	10.7	704	1	CRAR_MOUSE	P98064 mus musculu
79	79	10.5	275	1	IL2A_FELCA	P41690 felis silve
80	77.5	10.3	347	1	HPT_RAT	P06866 rattus norv
81	77.5	10.3	372	1	LEM1_MOUSE	P18337 mus musculu
82	76.5	10.2	699	1	CRAR_HUMAN	P48740 h complemen
83	76	10.1	1248	1	DIA1_HUMAN	O60610 homo sapien
84	75.5	10.0	1627	1	PAPA_HUMAN	Q13219 homo sapien
85	74.5	9.9	272	1	IL2A_HUMAN	P01589 homo sapien
86	74.5	9.9	705	1	C1R_HUMAN	P00736 homo sapien
87	72.5	9.6	1257	1	PGCN_RAT	P55067 rattus norv
88	72	9.6	347	1	HPT_ATEGE	P50417 ateles geof
89	72	9.6	745	1	ZP2_HUMAN	Q05996 homo sapien
90	72	9.6	960	1	GBR1_MOUSE	Q9WV18 mus musculu
91	72	9.6	961	1	GBR1_HUMAN	Q9UBS5 homo sapien
92	72	9.6	991	1	GBR1_RAT	Q9Z0U4 rattus norv
93	72	9.6	1255	1	DIA1_MOUSE	O08808 mus musculu
94	71	9.4	291	1	DM43_DIDMA	P82957 didelphis m
95	71	9.4	3381	1	PGCV_BOVIN	P81282 bos taurus
96	70.5	9.4	651	1	Y942_METJA	Q58852 methanococc
97	70.5	9.4	837	1	MUC1_RAT	P98089 rattus norv
98	70	9.3	666	1	ZP2_RABIT	P48829 oryctolagus
99	70	9.3	745	1	ZP2_MACFA	Q77725 macaca radi
100	70	9.3	839	1	PHLD_BOVIN	P80109 bos taurus

ALIGNMENTS

RESULT 1

```
CR2_MOUSE
ID   CR2_MOUSE      STANDARD;      PRT;  1025 AA.
AC   P19070;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   31-FEB-1996 (Rel. 34, Last annotation update)
DE   Complement receptor type 2 precursor (Cr2) (Complement C3d receptor).
GN   CR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=90229735; PubMed=2139457;
RA   Fingeroth J.D.;
RT   "Comparative structure and evolution of murine CR2. The homolog of
RT   the human C3d/EBV receptor (CD21).";
RL   J. Immunol. 144:3458-3467(1990).
RN   [2]
RP   SEQUENCE OF 12-1025 FROM N.A.
RX   MEDLINE=91010789; PubMed=2145366;
RA   Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT   "A molecular and immunochemical characterization of mouse CR2.
RT   Evidence for a single gene model of mouse complement receptors 1 and
RT   2.";
RL   J. Immunol. 145:2974-2983(1990).
RN   [3]
RP   SEQUENCE OF 344-401 AND 991-1025 FROM N.A.
RX   MEDLINE=89098890; PubMed=2784485;
RA   Fingeroth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT   "Identification of murine complement receptor type 2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN   [4]
RP   SEQUENCE OF 289-1025 FROM N.A.
RX   MEDLINE=89381450; PubMed=2528587;
RA   Kurtz C.B., Paul M.S., Aggerter M., Weis J.J., Weis J.H.;
RT   "Murine complement receptor gene family. II. Identification and
RT   characterization of the murine homolog (Cr2) to human CR2 and its
RT   molecular linkage to Crry.";
RL   J. Immunol. 143:2058-2067(1989).
CC   -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC       LYMPHOCYTES ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC   -!- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC   -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81083; AAA37451.1; -.
DR   EMBL; M35684; AAA37448.1; -.
DR   EMBL; M61132; AAA64295.1; -.
DR   EMBL; M35685; AAA37450.1; ALT_SEQ.
DR   EMBL; M29281; AAA37447.1; -.
DR   PIR; A43526; A43526.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88489; Cr2.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 14.
DR   SMART; SM00032; CCP; 14.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi.
FT   SIGNAL       1     11
FT   CHAIN       12   1025       COMPLEMENT RECEPTOR TYPE 2.
FT   DOMAIN      12    963       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    964    990       POTENTIAL.
FT   DOMAIN      991   1025       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       13     74       SUSHI 1.
FT   DOMAIN       81    139       SUSHI 2.
FT   DOMAIN      145    203       SUSHI 3.
FT   DOMAIN      206    264       SUSHI 4.
FT   DOMAIN      267    335       SUSHI 5.
FT   DOMAIN      342    399       SUSHI 6.
FT   DOMAIN      401    458       SUSHI 7.
FT   DOMAIN      462    515       SUSHI 8.
FT   DOMAIN      518    586       SUSHI 9.
FT   DOMAIN      593    649       SUSHI 10.
FT   DOMAIN      653    705       SUSHI 11.
FT   DOMAIN      708    769       SUSHI 12.
FT   DOMAIN      777    834       SUSHI 13.
FT   DOMAIN      840    898       SUSHI 14.
FT   DOMAIN      901    959       SUSHI 15.
FT   DISULFID     14     56       BY SIMILARITY.
FT   DISULFID     42     73       BY SIMILARITY.
FT   DISULFID     82    124       BY SIMILARITY.
FT   DISULFID    110    138       BY SIMILARITY.
FT   DISULFID    146    189       BY SIMILARITY.
FT   DISULFID    175    202       BY SIMILARITY.
FT   DISULFID    207    248       BY SIMILARITY.
FT   DISULFID    234    263       BY SIMILARITY.
FT   DISULFID    268    317       BY SIMILARITY.
FT   DISULFID    297    334       BY SIMILARITY.
FT   DISULFID    343    385       BY SIMILARITY.
FT   DISULFID    371    398       BY SIMILARITY.
FT   DISULFID    402    445       BY SIMILARITY.
FT   DISULFID    431    458       BY SIMILARITY.
FT   DISULFID    463    501       BY SIMILARITY.
FT   DISULFID    487    514       BY SIMILARITY.
FT   DISULFID    519    568       BY SIMILARITY.
FT   DISULFID    548    585       BY SIMILARITY.
FT   DISULFID    594    636       BY SIMILARITY.
FT   DISULFID    622    649       BY SIMILARITY.
FT   DISULFID    654    689       BY SIMILARITY.
FT   DISULFID    675    704       BY SIMILARITY.
FT   DISULFID    709    752       BY SIMILARITY.
FT   DISULFID    738    769       BY SIMILARITY.
FT   DISULFID    778    820       BY SIMILARITY.
FT   DISULFID    806    833       BY SIMILARITY.
FT   DISULFID    841    884       BY SIMILARITY.
FT   DISULFID    870    897       BY SIMILARITY.
FT   DISULFID    902    945       BY SIMILARITY.
FT   DISULFID    931    958       BY SIMILARITY.
FT   CARBOHYD     77     77       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    276    276       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    364    364       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    484    484       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    527    527       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    694    694       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    754    754       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    790    790       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    813    813       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    823    823       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    851    851       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    901    901       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    289    291       YGS -> EER (IN REF. 4).
FT   CONFLICT    306    306       S -> T (IN REF. 2).
FT   CONFLICT    520    520       P -> A (IN REF. 2).
FT   CONFLICT    962    963       MISSING (IN REF. 4).
SQ   SEQUENCE   1025 AA;  112994 MW;  19E51B09A0273694 CR0P4.

  Query Match          100.0%;  Score 752;  DB 1;  Length 1025;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-68;
  Matches  135;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 70

Qy      61 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 130

Qy     121 GPTALPVCESDFPLE 135
           |||||||||||||||
Db     131 GPTALPVCESDFPLE 145


RESULT 2
CR2_HUMAN
ID   CR2_HUMAN       STANDARD;       PRT:  1033 AA.
AC   P20023;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement receptor type 2 precursor (Cr2) (Complement C3d receptor)
DE   (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).
GN   CR2 OR C3DR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89123277; PubMed=2563370;
RA   Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA   Bolers V.M.;
RT   "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT   virus receptor.";
RL   J. Biol. Chem. 264:2118-2125(1989).
RN   [2]
RP   SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RX   MEDLINE=86287311; PubMed=3016712;
RA   Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA   de Bruyn Kops A., Smith J.A., Weis J.H.;
RT   "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT   virus receptor of human B lymphocytes: homology with the receptor for
RT   fragments C4b and C4b of the third and fourth components of
RT   complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
RN   [3]
RP   SEQUENCE OF 492-556 FROM N.A. (SHORT ISOFORM).
RX   MEDLINE=93294286; PubMed=8390533;
RA   Sinha S.K., Todd S.C., Hedrick J.A., Speiser C.L., Lambris J.D.,
RA   Tsoukas C.D.;
RT   "Characterization of the EBV/C4d receptor on the human Jurkat T cell
RT   line: evidence for a novel transcript.";
RL   J. Immunol. 150:5311-5320(1993).
CC   -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR
CC       VIRUS ON HUMAN B CELLS AND T-CELLS. PARTICIPATES IN B LYMPHOCYTES
CC       ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES, T LYMPHOCYTES AND
CC       FOLLICULAR DENDRITIC CELLS OF THE SPLEEN.
CC   -!- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA) FAMILY.
CC   -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD21 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd21.htm".
CC   ---------- ---------- ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ---------- ----------
DR   EMBL; M26004; AAA35786.1; -.
DR   EMBL; M26006; AAB04638.1; -.
DR   EMBL; M24007; AAB04638.1; JOINED.
DR   EMBL; M24008; AAB04638.1; JOINED.
DR   EMBL; M24009; AAB04638.1; JOINED.
DR   EMBL; M24010; AAB04638.1; JOINED.
DR   EMBL; M24011; AAB04638.1; JOINED.
DR   EMBL; M26009; AAB04638.1; JOINED.
DR   EMBL; M26010; AAB04638.1; JOINED.
DR   EMBL; M26011; AAB04638.1; JOINED.
DR   EMBL; M26012; AAB04638.1; JOINED.
DR   EMBL; M26013; AAB04638.1; JOINED.
DR   EMBL; M26014; AAB04638.1; JOINED.
DR   EMBL; M26015; AAB04638.1; JOINED.
DR   EMBL; S62696; AAB27186.1; -.
DR   PIR; A32036; A32036.
DR   PIR; A24319; A24319.
DR   PIR; B24319; B24319.
DR   PIR; C24319; C24319.
DR   PIR; D24319; D24319.
DR   PIR; E24319; E24319.
DR   PIR; F24319; F24319.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120650; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP
DR   Pfam; PF00084; sushi; 15.
DR   SMART; SM00032; CCP; 14.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi; Alternative splicing.
FT   SIGNAL       1     20
FT   CHAIN       21   1033       COMPLEMENT RECEPTOR TYPE 2.
FT   DOMAIN      21    971       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    972   999       POTENTIAL.
FT   DOMAIN     1000   1033      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      22     83       SUSHI 1.
FT   DOMAIN      90    147       SUSHI 2.
FT   DOMAIN     153    211       SUSHI 3.
FT   DOMAIN     214    272       SUSHI 4.
FT   DOMAIN     275    343       SUSHI 5.
FT   DOMAIN     350    407       SUSHI 6.
FT   DOMAIN     409    467       SUSHI 7.
FT   DOMAIN     470    523       SUSHI 8.
FT   DOMAIN     526    594       SUSHI 9.
FT   DOMAIN     601    658       SUSHI 10.
FT   DOMAIN     661    715       SUSHI 11.
FT   DOMAIN     718    780       SUSHI 12.
FT   DOMAIN     787    844       SUSHI 13.
FT   DOMAIN     850    908       SUSHI 14.
FT   DOMAIN     911    969       SUSHI 15.
FT   DISULFID    23     65       BY SIMILARITY.
FT   DISULFID    51     82       BY SIMILARITY.
FT   DISULFID    91    132       BY SIMILARITY.
FT   DISULFID   118    146       BY SIMILARITY.
FT   DISULFID   154    197       BY SIMILARITY.
FT   DISULFID   183    210       BY SIMILARITY.
FT   DISULFID   215    256       BY SIMILARITY.
FT   DISULFID   242    271       BY SIMILARITY.
FT   DISULFID   276    325       BY SIMILARITY.
FT   DISULFID   305    342       BY SIMILARITY.
FT   DISULFID   351    393       BY SIMILARITY.
FT   DISULFID   379    406       BY SIMILARITY.
FT   DISULFID   410    453       BY SIMILARITY.
FT   DISULFID   439    466       BY SIMILARITY.
FT   DISULFID   471    509       BY SIMILARITY.
FT   DISULFID   495    522       BY SIMILARITY.
FT   DISULFID   527    576       BY SIMILARITY.
FT   DISULFID   556    593       BY SIMILARITY.
FT   DISULFID   602    644       BY SIMILARITY.
FT   DISULFID   630    657       BY SIMILARITY.
FT   DISULFID   662    699       BY SIMILARITY.
```

```
FT   DISULFID     685     714        BY SIMILARITY.
FT   DISULFID     719     762        BY SIMILARITY.
FT   DISULFID     738     779        BY SIMILARITY.
FT   DISULFID     788     840        BY SIMILARITY.
FT   DISULFID     816     843        BY SIMILARITY.
FT   DISULFID     851     894        BY SIMILARITY.
FT   DISULFID     880     907        BY SIMILARITY.
FT   DISULFID     912     955        BY SIMILARITY.
FT   DISULFID     941     968        BY SIMILARITY.
FT   CARBOHYD     121     121        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     127     127        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     294     294        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     372     372        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     492     492        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     623     623        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     682     682        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     800     800        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     823     823        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     861     861        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     911     911        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     499     524        MISSING (IN SHORT ISOFORM).
FT   VARSPLIC     525     556        IFCPPPPPVIYNGAHTGSSLEDFPYGTIVTYTC - - NHLPT
FT                                   IPCYLQWGTHREFLRRFSIWNHGHLHM (IN SHORT
FT                                   ISOFORM).
FT   CONFLICT     667     667        Q - D (IN REF. 2).
FT   CONFLICT     902     902        Q - G (IN REF. 2).
FT   CONFLICT     906     906        R - L (IN REF. 2).
SQ   SEQUENCE   1033 AA;  112973 MW;  1749DB4A07847ADA CRC64;

  Query Match             62.0%;  Score 466.5;  DB 1;  Length 1033;
  Best Local Similarity   61.2%;  Pred. No. 4e-49;
  Matches   82;  Conservative   18;  Mismatches   33;  Indels    1;  Gaps    1;

Qy        2 ISCDPPPEVKNARKPYYSLEPTVPCTVGRYTCSPSYRLIGEKATFCTSENQVHATWDKAPP 61
            ||| || | | | |  |:| ||  |||:|||:| ::|||||:: ||:::|  |||| |
Db       21 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGTFRLIGEKSILCITKDKVDGTWDKPAP 80

Qy       62 TCESVNKPISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQANEMWG 121
            ||  || | 3| :|||:||:  :|| |:|||||||| || 3|:| |:|:|||||| |||
Db       81 KCEYSNKYSSCSEPLVPGGYKIKGS IPYRHGDSVIFACKTNFSMNGNKSVWCQANNMWG 139

Qy      122 PTALPVCESDFPLE 135
            || || | | ||||
Db      140 PTRLPTCVSVFPLE 153


RESULT 3
C4BP_HUMAN
ID   C4BP_HUMAN      STANDARD;      PRT;   597 AA.
AC   P04003;
DT   23-OCT-1986 (Rel. 02, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein alpha chain precursor (C4bp) (Proline-rich
DE   protein) (PRP).
GN   C4BPA OR C4BP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90073699; PubMed=2596215;
RA   Matsushuchi F., Okamura S., Aso T., Sata T., Niho Y.;
RT   "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT   identity of PRP as C4b-binding protein.";
RL   Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91113199; PubMed=1989602;
RA   Aso T., Okamura S., Matsushuchi F., Sakamoto N., Sata T., Niho Y.;
RT   "Genomic organization of the alpha-chain of the human C4b-binding
RT   protein gene.";
RL   Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN   [3]
RP   SEQUENCE OF 9-81 FROM N.A.
RX   MEDLINE=88242821; PubMed=3378624;
RA   Lintin S.J., Lowin A.R., Reid K.B.M.;
RT   "Derivation of the sequence of the signal peptide in human
RT   C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT   cDNA sequence.";
RL   FEBS Lett. 232:328-332(1988).
RN   [4]
RP   SEQUENCE OF 204-288 FROM N.A.
RX   MEDLINE=86401119; PubMed=3017751;
RA   Lintin S.J., Reid K.B.M.;
RT   "Studies on the structure of the human C4b binding protein gene.";
RL   FEBS Lett. 204:77-81(1986).
RN   [5]
RP   SEQUENCE OF 80-597 FROM N.A.
RX   MEDLINE=86025405; PubMed=3840370;
RA   Chung L.P., Bentley D.R., Reid K.B.M.;
RT   "Molecular cloning and characterization of the cDNA coding for C4b
RT   binding protein, a regulatory protein of the classical pathway of the
RT   human complement system.";
RL   Biochem. J. 230:133-141(1985).
RN   [6]
RP   SEQUENCE OF 49-88.
RX   MEDLINE=85296001; PubMed=4033666;
RA   Chung L.P., Gagnon J., Reid K.B.M.;
RT   "Amino acid sequence studies of human C4b-binding protein: N-terminal
RT   sequence analysis and alignment of the fragments produced by limited
RT   proteolysis with chymotrypsin and the peptides produced by cyanogen
RT   bromide treatment.";
RL   Mol. Immunol. 22:427-435(1985).
RN   [7]
RP   ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND-BINDING.
RX   MEDLINE=87221615; PubMed=6222381;
RA   Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT   "Visualization of human C4b-binding protein and its complexes with
RT   vitamin K-dependent protein S and complement protein C4b.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C4B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC       AND WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC       OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC       BETA CHAIN, A 540 kDa HOMOHEPTAMER OF ALPHA CHAINS OR A 500 kDa
CC       COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC       THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC       BINDING SITE FOR C4B AT THE END.
CC   -!- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC   ---------- - ------------ -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- - ------------ -----
DR   EMBL; M31452; AAA36507.1; -.
DR   EMBL; M62486; AAA36506.1; -.
DR   EMBL; M62475; AAA36506.1; JOINED.
DR   EMBL; M62476; AAA36506.1; JOINED.
DR   EMBL; M62477; AAA36506.1; JOINED.
DR   EMBL; M62478; AAA36506.1; JOINED.
DR   EMBL; M62479; AAA36506.1; JOINED.
```

```
DR   EMBL; M62480; AAA36506.1; JOINED.
DR   EMBL; M62481; AAA36506.1; JOINED.
DR   EMBL; M62482; AAA36506.1; JOINED.
DR   EMBL; M62484; AAA36506.1; JOINED.
DR   EMBL; M62485; AAA36506.1; JOINED.
DR   EMBL; X07854; CAA30701.1; -.
DR   EMBL; X04284; CAB51244.1; -.
DR   EMBL; X04296; CAA27849.1; -.
DR   EMBL; X02865; CAA26617.1; -.
DR   PIR; A34568; NBHUC4.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120830; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW   Polymorphism.
FT   SIGNAL        1     48
FT   CHAIN        49    597       C4B-BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN       49    109       SUSHI 1.
FT   DOMAIN      112    171       SUSHI 2.
FT   DOMAIN      174    235       SUSHI 3.
FT   DOMAIN      238    295       SUSHI 4.
FT   DOMAIN      298    361       SUSHI 5.
FT   DOMAIN      364    423       SUSHI 6.
FT   DOMAIN      425    481       SUSHI 7.
FT   DOMAIN      483    539       SUSHI 8.
FT   DISULFID     50     96       BY SIMILARITY.
FT   DISULFID     81    108       BY SIMILARITY.
FT   DISULFID    113    154       BY SIMILARITY.
FT   DISULFID    140    170       BY SIMILARITY.
FT   DISULFID    175    217       BY SIMILARITY.
FT   DISULFID    203    234       BY SIMILARITY.
FT   DISULFID    239    281       BY SIMILARITY.
FT   DISULFID    267    294       BY SIMILARITY.
FT   DISULFID    299    348       BY SIMILARITY.
FT   DISULFID    332    360       BY SIMILARITY.
FT   DISULFID   2364   2387       BY SIMILARITY.
FT   DISULFID   2365   2409       BY SIMILARITY.
FT   DISULFID   2399    422       BY SIMILARITY.
FT   DISULFID    426    468       BY SIMILARITY.
FT   DISULFID    454    480       BY SIMILARITY.
FT   DISULFID    484    525       BY SIMILARITY.
FT   DISULFID    511    538       BY SIMILARITY.
FT   DISULFID    546    546       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID    558    558       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   CARBOHYD    221    221       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    506    506       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .).
FT   VARIANT      92     92       Q -> L.
FT                                /FTId=VAR_001977.
FT   VARIANT     357    357       Y -> H.
FT                                /FTId=VAR_001978.
FT   VARIANT     473    473       W -> L (IN DBSNP:1801341).
FT                                /FTId=VAR_012038.
SQ   SEQUENCE   597 AA;  67033 MW;  67E03F2EA85A16DD CRC64;

  Query Match             40.0%;  Score 225.5;  DB 1;  Length 597;
  Best Local Similarity   41.8%;  Pred. No. 3.4e-15;
  Matches   42;  Conservative   25;  Mismatches   52;  Indels   13;  Gaps    4;

Qy       2 ISCDPPPEVKNAR    KPYYSLPIVPGTVERYTCSPSYRLIGEKAIFCISENQVHATWD 57
           : | |||:::  |    : :|:    | : |:| | : |:|  :| |  ||:   |
Db     173 VKCKPPPDIRNGRHSGEENFYAY-   GFSVTYSCDPRFSLLGHASISCTVENETIGVWR 228

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
           :|| ||    |:|  | | | | |    : : |:: | |:  | ::||  : | |:
Db     229 PSPPTCEK-  -ITCRKPDVSHGEMVSGFGPIYNYKDTIVFKCQKGFVLRGSSVIHCDAD 284

Qy     118 EMWGPTALPVCE 129
             | |:  | :|
Db     285 SKWNPSP PACE 295
```

```
RESULT 4
C4BP_MOUSE
ID   C4BP_MOUSE      STANDARD;       PRT:   469 AA.
AC   P08607;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b binding protein precursor (C4bp)
GN   C4BPA OR C4BP
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88024997; PubMed=3663616;
RA   Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT   "cDNA structure of murine C4b-binding protein, a regulatory component
RT   of the serum complement system.";
RL   Biochemistry 26:4668-4674(1987).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA) WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: HOMOHEPTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
CC       BETA CHAIN OF C4BP.
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET1 OR MET 44 IS THE INITIATOR.
CC   ---------- ---------- ... ... ----------  ..
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------- ---------- ... ... ----------  ..
DR   EMBL; M17122; AAA37312.1; ALT_INIT.
DR   PIR; A27117; NBMSC4.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:88279; C4bp.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1     56
FT   CHAIN        57    469       C4B-BINDING PROTEIN.
FT   DOMAIN       57    116       SUSHI 1.
FT   DOMAIN      119    177       SUSHI 2.
FT   DOMAIN      180    241       SUSHI 3.
FT   DOMAIN      244    300       SUSHI 4.
FT   DOMAIN      302    356       SUSHI 5.
FT   DOMAIN      358    414       SUSHI 6.
FT   DISULFID     58    103       BY SIMILARITY.
FT   DISULFID     88    115       BY SIMILARITY.
FT   DISULFID    120    160       BY SIMILARITY.
FT   DISULFID    146    176       BY SIMILARITY.
FT   DISULFID    181    223       BY SIMILARITY.
FT   DISULFID    209    240       BY SIMILARITY.
FT   DISULFID    245    287       BY SIMILARITY.
FT   DISULFID    273    299       BY SIMILARITY.
FT   DISULFID    303    343       BY SIMILARITY.
FT   DISULFID    329    355       BY SIMILARITY.
FT   DISULFID    359    400       BY SIMILARITY.
FT   DISULFID    386    413       BY SIMILARITY.
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    127    227       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    275    275       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD     66     66       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    428    428       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   469 AA;  51561 MW;  41E1437CB8DBC6321 CRC64;

  Query Match             29.5%;  Score 222;  DB 1;  Length 469;
  Best Local Similarity   64.3%;  Pred. No. 5.8e-15;
  Matches   47;  Conservative   17;  Mismatches   57;  Indels   11;  Gaps    5;

Qy       2 ISCDPPPEVKNAR    KFYYSLFTVPGIVLRYTCSPSYRLIGEKAIPCISENQVHATWD 57
           : : |||:  | :    : :|  |   : ||| | :|||:   | |   |:    |
Db     176 VKCGPPPDIDSNGKHSGIEFEY   PYNHC   ISYTCDPGFRLVGSPFIQCTVVNKTVPVWS 234

Qy      58 KAPPICESVNKTISCSDITVDGGFMNKGSKAPFRHGDSVTFICKANFTMKGSKIVWCQAN 117
           :|. ||     | || : :  : :  4 || : | |||   |     ::|   : || |
Db     235 SSPPTCEK    IICSQPNITHGVIVSGYKATYTHRDSVRLACLNGTVLRGRHVIECQGN 290

Qy     118 EMWGPTALPVCESDEPL 134
             1   :|: :| | |  |
Db     291 GNW  SSLPICEEVDCL 307


RESULT 5
CR1_HUMAN
ID   CR1_HUMAN       STANDARD;       PRT;  2039 AA.
AC   P17927;
DT   01-NOV-1990 (Rel. 18, Created)
DT   01-NOV-1990 (Rel. 18, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
DE   antigen).
GN   CR1 OR C3BR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89043992; PubMed=2972794;
RA   Klickstein L.B., Barlow P.J., Miletic V., Rabson L.D., Smith J.A.,
RA   Fearon D.T.;
RT   "Identification of distinct C3b and C4b recognition sites in the
RT   human C3b/C4b receptor (CR1, CD35) by deletion mutagenesis.";
RL   J. Exp. Med. 168:1699-1717(1988).
RN   [2]
RP   SEQUENCE OF 508-2039 FROM N.A.
RX   MEDLINE=87168191; PubMed=2951476;
RA   Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA   Fearon D.T.;
RT   "Human C3b/C4b receptor (CR1). Demonstration of long homologous
RT   repeating domains that are composed of the short consensus repeats
RT   characteristic of C3/C4 binding proteins.";
RL   J. Exp. Med. 165:1095-1112(1987).
RN   [3]
RP   SEQUENCE OF 761-784; 831-854 AND 1179-1195 FROM N.A.
RX   MEDLINE=86067975; PubMed=2934745;
RA   Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Pearon D.T.;
RT   "Identification of a partial cDNA clone for the human receptor for
RT   complement fragments C3b/C4b.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC   -!- FUNCTION: CR1, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC       PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS, MEDIATES THE
CC       BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC       ACTIVATED COMPLEMENT.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- POLYMORPHISM: CR1 IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.
CC   -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC       LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SCRS OF LHR A
CC       CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC       TWO SCRS OF LHR B AND C EACH HAD A SITE DETERMINING C3
CC       SPECIFICITY.
CC   -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CR1.
CC   -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC   ---------- ... ---------- ... ---------- ... ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ... ---------- ... ---------- ... ----------
DR   EMBL; M11569; AAA52297.1; -.
DR   EMBL; M11673; AAA52298.1; -.
DR   EMBL; M11678; AAA52299.1; -.
DR   EMBL; Y00845; CAA68755.1; -.
DR   EMBL; X05430; CAA28933.1; -.
DR   PIR; A28507; A28507.
DR   PIR; A24748; A24748.
DR   PIR; B24748; B24748.
DR   PIR; C24748; C24748.
DR   PIR; S04843; S04843.
DR   HSSP; P08603; 1HFI.
DR   MIM; 120620; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi; Blood group antigen.
FT   SIGNAL        1     41
FT   CHAIN        42   2039       COMPLEMENT RECEPTOR TYPE 1.
FT   DOMAIN       42   1971       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1972   1996       POTENTIAL.
FT   DOMAIN     1997   2039       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES      42     42       PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT   DOMAIN       42    100       SUSHI A1.
FT   DOMAIN      103    162       SUSHI A2.
FT   DOMAIN      165    233       SUSHI A3.
FT   DOMAIN      237    294       SUSHI A4.
FT   DOMAIN      296    354       SUSHI A5.
FT   DOMAIN      357    417       SUSHI A6.
FT   DOMAIN      420    488       SUSHI A7.
FT   DOMAIN      492    550       SUSHI B1.
FT   DOMAIN      553    612       SUSHI B2.
FT   DOMAIN      615    683       SUSHI B3.
FT   DOMAIN      687    744       SUSHI B4.
FT   DOMAIN      746    804       SUSHI B5.
FT   DOMAIN      807    867       SUSHI B6.
FT   DOMAIN      870    938       SUSHI B7.
FT   DOMAIN      942   1000       SUSHI C1.
FT   DOMAIN     1003   1062       SUSHI C2.
FT   DOMAIN     1065   1133       SUSHI C3.
FT   DOMAIN     1137   1194       SUSHI C4.
FT   DOMAIN     1196   1254       SUSHI C5.
FT   DOMAIN     1257   1317       SUSHI C6.
FT   DOMAIN     1320   1388       SUSHI C7.
FT   DOMAIN     1395   1453       SUSHI D1.
FT   DOMAIN     1456   1515       SUSHI D2.
FT   DOMAIN     1518   1586       SUSHI D3.
FT   DOMAIN     1590   1647       SUSHI D4.
FT   DOMAIN     1649   1707       SUSHI D5.
FT   DOMAIN     1710   1770       SUSHI D6.
FT   DOMAIN     1773   1841       SUSHI D7.
FT   DOMAIN     1847   1905       SUSHI E1.
FT   DOMAIN     1908   1966       SUSHI E2.
FT   DISULFID     43     86       BY SIMILARITY.
FT   DISULFID     73     99       BY SIMILARITY.
FT   DISULFID    104    145       BY SIMILARITY.
FT   DISULFID    131    161       BY SIMILARITY.
FT   DISULFID    166    215       BY SIMILARITY.
FT   DISULFID    195    232       BY SIMILARITY.
FT   DISULFID    238    280       BY SIMILARITY.
FT   DISULFID    266    293       BY SIMILARITY.
FT   DISULFID    297    340       BY SIMILARITY.
FT   DISULFID    326    353       BY SIMILARITY.
```

```
FT   DISULFID    358    400       BY SIMILARITY.
FT   DISULFID    386    416       BY SIMILARITY.
FT   DISULFID    421    470       BY SIMILARITY.
FT   DISULFID    450    487       BY SIMILARITY.
FT   DISULFID    493    536       BY SIMILARITY.
FT   DISULFID    523    549       BY SIMILARITY.
FT   DISULFID    554    595       BY SIMILARITY.
FT   DISULFID    581    611       BY SIMILARITY.
FT   DISULFID    615    665       BY SIMILARITY.
FT   DISULFID    645    682       BY SIMILARITY.
FT   DISULFID    688    740       BY SIMILARITY.
FT   DISULFID    716    743       BY SIMILARITY.
FT   DISULFID    747    790       BY SIMILARITY.
FT   DISULFID    776    803       BY SIMILARITY.
FT   DISULFID    808    850       BY SIMILARITY.
FT   DISULFID    836    866       BY SIMILARITY.
FT   DISULFID    871    920       BY SIMILARITY.
FT   DISULFID    903    937       BY SIMILARITY.
FT   DISULFID    943    986       BY SIMILARITY.
FT   DISULFID    973    999       BY SIMILARITY.
FT   DISULFID   1004   1045       BY SIMILARITY.
FT   DISULFID   1031   1061       BY SIMILARITY.
FT   DISULFID   1066   1115       BY SIMILARITY.
FT   DISULFID   1095   1132       BY SIMILARITY.
FT   DISULFID   1138   1180       BY SIMILARITY.
FT   DISULFID   1166   1193       BY SIMILARITY.
FT   DISULFID   1197   1240       BY SIMILARITY.
FT   DISULFID   1226   1253       BY SIMILARITY.
FT   DISULFID   1258   1300       BY SIMILARITY.
FT   DISULFID   1286   1316       BY SIMILARITY.
FT   DISULFID   1321   1370       BY SIMILARITY.
FT   DISULFID   1350   1387       BY SIMILARITY.
FT   DISULFID   1396   1439       BY SIMILARITY.
FT   DISULFID   1426   1452       BY SIMILARITY.
FT   DISULFID   1457   1498       BY SIMILARITY.
FT   DISULFID   1484   1514       BY SIMILARITY.
FT   DISULFID   1519   1568       BY SIMILARITY.
FT   DISULFID   1548   1585       BY SIMILARITY.
FT   DISULFID   1591   1633       BY SIMILARITY.
FT   DISULFID   1619   1646       BY SIMILARITY.
FT   DISULFID   1650   1693       BY SIMILARITY.
FT   DISULFID   1679   1706       BY SIMILARITY.
FT   DISULFID   1713   1753       BY SIMILARITY.
FT   DISULFID   1739   1769       BY SIMILARITY.
FT   DISULFID   1774   1823       BY SIMILARITY.
FT   DISULFID   1803   1840       BY SIMILARITY.
FT   DISULFID   1848   1891       BY SIMILARITY.
FT   DISULFID   1877   1904       BY SIMILARITY.
FT   DISULFID   1909   1952       BY SIMILARITY.
FT   DISULFID   1938   1965       BY SIMILARITY.
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    252    252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    447    447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    860    860       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    897    897       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    959    959       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1028   1028       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1152   1152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1310   1310       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1481   1481       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1504   1504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1534   1534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1540   1540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1605   1605       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1763   1763       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1908   1908       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2039 AA;  223589 MW;  B2FD29B6AD3C5EB7 CRC64;

  Query Match          29.3%;  Score 220.5;  DB 1;  Length 2039;
  Best Local Similarity   36.0%;  Pred. No. 4.1e-14;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS---  --YRLIGEKAIFCISENQV 52
           |||:|||: : |    :||        |||: | |         : |:|||:|:| |::
Db    1537 ISCEPPPTISNG--DFYSNNRTSFHNGTVVIYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1594

Qy      53 HAIWDKAPLICESVNKFISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKIV 112
           |   || | | ||   |: | |       |::: |   : : | |:  | | || |:
Db    1595 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 1651

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1652 QCQTNGRWGP-KLPHC 1667


RESULT 6
C4BP_RAT
ID   C4BP_RAT       STANDARD;      PRT:   558 AA.
AC   Q63514;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b binding protein alpha chain precursor (C4bp).
GN   C4BPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=97166082; PubMed=9013975;
RA   Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT   "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
RT   structural and functional relationships among human, bovine, rabbit,
RT   mouse, and rat proteins.";
RL   J. Immunol. 158:1315-1323(1997).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC       AND WITH SERUM AMYLOID P COMPONENT
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; Z50051; CAA90391.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1     13       BY SIMILARITY.
FT   CHAIN        14    558       C4B-BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN       14     73       SUSHI 1.
FT   DOMAIN       76    135       SUSHI 2.
FT   DOMAIN      138    200       SUSHI 3.
FT   DOMAIN      203    259       SUSHI 4.
FT   DOMAIN      262    325       SUSHI 5.
FT   DOMAIN      328    387       SUSHI 6.
FT   DOMAIN      389    444       SUSHI 7.
FT   DOMAIN      446    502       SUSHI 8.
```

```
FT   DISULFID     15     60       BY SIMILARITY.
FT   DISULFID     45     72       BY SIMILARITY.
FT   DISULFID     77    118       BY SIMILARITY.
FT   DISULFID    104    134       BY SIMILARITY.
FT   DISULFID    149    182       BY SIMILARITY.
FT   DISULFID    198    199       BY SIMILARITY.
FT   DISULFID    204    246       BY SIMILARITY.
FT   DISULFID    242    258       BY SIMILARITY.
FT   DISULFID    263    312       BY SIMILARITY
FT   DISULFID    296    424       BY SIMILARITY.
FT   DISULFID   2428   2491       BY SIMILARITY.
FT   DISULFID   2429   2473       BY SIMILARITY.
FT   DISULFID   2463    886       BY SIMILARITY.
FT   DISULFID    390    431       BY SIMILARITY.
FT   DISULFID    417    443       BY SIMILARITY.
FT   DISULFID    447    488       BY SIMILARITY.
FT   DISULFID    474    501       BY SIMILARITY.
FT   DISULFID    509    509       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID    521    521       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    177    177       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    186    186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    469    469       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    491    491       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   558 AA;  62269 MW;  592F0C667ED1E5FF CRC64;

  Query Match             27.7%;  Score 208.5;  DB 1;  Length 558;
  Best Local Similarity   30.6%;  Pred. No. 1.6e-13;
  Matches   44;  Conservative   26;  Mismatches   51;  Indels   19;  Gaps    5;

Qy        4 CDPPPEVKNAR      KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDK 58
            |  ||:: : :      : :::       : : | | | : |:|  :| |   |:    |
Db      139 CGMPPDISNGKHNGREEEFFYY     RSSVTYKCDPDFTLIGNASITCTVVNKTVGVWSP 194

Qy       59 APPICPSVNKTISCSDPIVPGGFMNKGSKAPFHRGDSVIFICKANFTMKGSKTVWCQANE 118
            :|| ||      | |  | | :  : | | |  ::: ||| | |:  | :::|  : |:|:
Db      195 SPPTCP     KLTCFWDKVIGRDIINSGFKHIYKYKDSVREVCQKGFVLRGSGVIHCEADG 250

Qy      119 MWGPTALPVCE     SDFP 133
              | |  :||||      :| |
Db      251 SWSP  VPVCELNSCTDIP 267


RESULT 7
C4BP_BOVIN
ID   C4BP_BOVIN      STANDARD;      PRT;   610 AA.
AC   Q28065;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein alpha chain precursor (C4bp).
GN   C4BPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95015009; PubMed=7590621;
RA   Hillarp A., Ekman A., Dahlback B.;
RT   "Bovine C4b binding protein. Molecular cloning of the alpha- and
RT   beta chains provides structural background for lack of complex
RT   formation with protein S.";
RL   J. Immunol. 153:4190-4199(1994).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C3/C4INA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DECAY OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC       CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z31695; CAA84498.1;  -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL       1     48       BY SIMILARITY.
FT   CHAIN       49    610       C4B-BINDING PROTEIN ALPHA CHAIN.
FT   DOMAIN      49    108       SUSHI 1.
FT   DOMAIN     111    170       SUSHI 2.
FT   DOMAIN     173    235       SUSHI 3.
FT   DOMAIN     238    295       SUSHI 4.
FT   DOMAIN     298    363       SUSHI 5.
FT   DOMAIN     366    426       SUSHI 6.
FT   DOMAIN     428    484       SUSHI 7.
FT   DOMAIN     486    542       SUSHI 8.
FT   DISULFID    50     95       BY SIMILARITY.
FT   DISULFID    80    107       BY SIMILARITY.
FT   DISULFID   112    153       BY SIMILARITY.
FT   DISULFID   139    169       BY SIMILARITY.
FT   DISULFID   174    217       BY SIMILARITY.
FT   DISULFID   203    234       BY SIMILARITY.
FT   DISULFID   239    281       BY SIMILARITY.
FT   DISULFID   267    294       BY SIMILARITY.
FT   DISULFID   299    350       BY SIMILARITY.
FT   DISULFID   334    362       BY SIMILARITY.
FT   DISULFID  ?366   ?390       BY SIMILARITY.
FT   DISULFID  ?367   ?412       BY SIMILARITY.
FT   DISULFID  ?402    425       BY SIMILARITY.
FT   DISULFID   429    471       BY SIMILARITY.
FT   DISULFID   457    483       BY SIMILARITY.
FT   DISULFID   487    528       BY SIMILARITY.
FT   DISULFID   514    541       BY SIMILARITY.
FT   DISULFID   549    549       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   DISULFID   561    561       INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT   CARBOHYD    66     66       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   221    221       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   525    525       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   602    602       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   610 AA;  68886 MW;  DB06B270EBA06B58 CRC64;

  Query Match             26.6%;  Score 200;  DB 1;  Length 610;
  Best Local Similarity   27.5%;  Pred. No. 1.2e-12;
  Matches   36;  Conservative   27;  Mismatches   54;  Indels   14;  Gaps    4;

Qy        4 CDPPPEVKNAR   ---KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDK 58
            |:||| : | |      : :|:     |: : |:|   : :::|: : /  |:    |
Db      174 CEPPGIISNGRHNGGDEDFYIY ---GSSVIYSCDRDFSMLGKASISCRVENKTIGVWSP 229

Qy       59 APPICPSVNKTISCSDPIVPGGFMNKGSKAPFHRGDSVIFICKANFTMKGSKTVWCQANE 118
            :|| |     | : |  |:|  | :  :  |    : :  |: : |    |::|  : |:|:
Db      230 SPPSC----KKVICVQPVVFDGRITSGFCPIYTYQQSIVYACNFCEMLGDSLIECEAEN 285

Qy      119 MWGPTALPVCE 129
              |  | ||
Db      286 SWNPPP-PICE 295


RESULT 8
VCP_VACCV
ID   VCP_VACCV       STANDARD;      PRT;   263 AA.
```

```
AC   P10998;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement control protein precursor (VCP) (Secretory protein 35)
DE   (Protein C4) (28 kDa protein).
GN   C3L.
OS   Vaccinia virus (strain WR), and
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254, 10249;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC   STRAIN=WR;
RX   MEDLINE=88318974; PubMed=3412473;
RA   Kotwal G.J., Moss B.;
RT   "Vaccinia virus encodes a secretory polypeptide structurally related
RT   to complement control proteins.";
RL   Nature 335:176-178(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WR;
RX   MEDLINE=89073756; PubMed=2849238;
RA   Kotwal G.J., Moss B.;
RT   "Analysis of a large cluster of nonessential genes deleted from a
RT   vaccinia virus terminal transposition mutant.";
RL   Virology 167:524-537(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COPENHAGEN;
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [4]
RP   COMPLETE GENOME.
RC   STRAIN=COPENHAGEN;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=92115714; PubMed=1731333;
RA   Isaacs S.N., Kotwal G.J., Moss B.;
RT   "Vaccinia virus complement-control protein prevents
RT   antibody-dependent complement-enhanced neutralization of infectivity
RT   and contributes to virulence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN   [6]
RP   STRUCTURE BY NMR OF 146-263.
RX   MEDLINE=97446168; PubMed=9299352;
RA   Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
RA   Barlow P.N.;
RT   "NMR studies of a viral protein that mimics the regulators of
RT   complement activation.";
RL   J. Mol. Biol. 272:253-265(1997).
CC   -!- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
CC       INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
CC       ACTIVATION. BINDS C3B AND C4B.
CC   -!- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13196; CAA31564.1; -.
DR   EMBL; M22812; AAA69605.1; -.
DR   EMBL; M35027; AAA47997.1; -.
DR   PIR; A31005; WMVZSP.
DR   PDB; 1VVC; 03-DEC-97.
DR   PDB; 1VVD; 03-DEC-97.
DR   PDB; 1VVE; 03-DEC-97.
DR   InterPro; IPR000436; Sushi_SCR_CCP
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Signal; Repeat; Sushi; 3D-structure.
FT   SIGNAL       1     19
FT   CHAIN       20    263       COMPLEMENT CONTROL PROTEIN.
FT   DOMAIN      20     82       SUSHI 1.
FT   DOMAIN      85    144       SUSHI 2.
FT   DOMAIN     147    202       SUSHI 3.
FT   DOMAIN     205    262       SUSHI 4.
FT   DISULFID    21     70       BY SIMILARITY.
FT   DISULFID    54     81       BY SIMILARITY.
FT   DISULFID    86    126       BY SIMILARITY.
FT   DISULFID   112    143       BY SIMILARITY.
FT   DISULFID   148    190       BY SIMILARITY.
FT   DISULFID   176    201       BY SIMILARITY.
FT   DISULFID   208    248       BY SIMILARITY.
FT   DISULFID   234    261       BY SIMILARITY.
SQ   SEQUENCE   263 AA;  28629 MW;  E432CC9A6EF8097 CRC64;

  Query Match             25.7%;  Score 19.5;  DB 1;  Length 263;
  Best Local Similarity   29.1%;  Pred. No. 2.2e-12;
  Matches   37;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVIRYTCSPSYPLIGEKAIFCISENQVHATWDFAPP 61
           :|  || : | |   |       |:|: | |:  | |||   : |             ||
Db     146 VKCQSPPSISNGRHNGYEDFYTDGSVVTYSCNPGYSLIGNSGVLCSG---GEWSD-PP 199

Qy      62 ICESVNKCISCSDPIVPGGFMNKGSKAPFRHCISVTFICKANFTMKGSKTVWCQANEMWG 121
           |:     : |  | :  |::: | |  : : |:| | ||  : : || :  |     |
Db     200 TCQ----CVKCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWK 255

Qy     122 PTALPVC 128
           |  || |
Db     256 PSELPKC 261


RESULT 9
MCP_HUMAN
ID   MCP_HUMAN      STANDARD;      PRT;   377 AA.
AC   P15529;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Membrane cofactor protein precursor (CD46 antigen) (Trophoblast
DE   leucocyte common antigen) (TLX).
GN   MCP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-58.
RX   MEDLINE=88286080; PubMed=3260937;
RA   Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,
RA   Rebentisch M.B., Lemons R.S., Seya T., Atkinson J.P.;
RT   "Molecular cloning and chromosomal localization of human membrane
RT   cofactor protein (MCP). Evidence for inclusion in the multigene
RT   family of complement-regulatory proteins.";
RL   J. Exp. Med. 168:181-194(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=93119658; PubMed=8418811;
RA   Cervoni F., Henichel P., Akhoundi S., Hsi B.L., Rossi B.;
```

```
RT   "Characterization of a cDNA clone coding for human testis membrane
RT   cofactor protein (MCP; CD46).";
RL   Mol. Reprod. Dev. 34:107-114(1993).
RN   [3]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91267562; PubMed=2050389;
RA   Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
RA   McKenzie I.F.;
RT   "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
RT   regulator of complement activation.";
RL   Immunogenetics 33:335-344(1991).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=94014456; PubMed=7691939;
RA   Ong W., Bourcade D., Post T., Greenlund A.C., Atkinson J.P.,
RA   Kumar V.;
RT   "Characterization of the promoter region of the membrane cofactor
RT   protein (CD46) gene of the human complement system and comparison to
RT   a membrane cofactor protein like genetic element.";
RL   J. Immunol. 151:4437-4446(1993).
RN   [5]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=92289809; PubMed=1601037;
RA   Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
RT   "Tissue-specific and allelic expression of the complement regulator
RT   CD46 is controlled by alternative splicing.";
RL   Eur. J. Immunol. 22:1513-1518(1992).
CC   -!- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
CC       CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
CC       FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
CC       ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
CC       CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
CC       THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
CC       ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
CC       SYNCYTIOTROPHOBLAST LAYER OF PLACENTA.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
CC       GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
CC       EBV-B CELLS AND LEUKEMIC CELLS. THE 66 kDa ALPHA ISOFORM AND THE
CC       56 kDa BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
CC       TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
CC       RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
CC       PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
CC       AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
CC       12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
CC   -!- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
CC       ERYTHROCYTES AND SOME BONE MARROW CELLS.
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
CC       (PROBABLE).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- DATABASE: NAME=MROW; NOTE=CD quride CD46 entry;
CC       WWW="http://www.ucl.ac.uk/mrow-cgi/genedb/CD46.html"
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensed@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00051; CAA68675.1; -.
DR   EMBL; S51940; AAB24802.1; -.
DR   EMBL; M58050; AAA62833.1; -.
DR   EMBL; A18585; CAA01400.1; -.
DR   EMBL; S65879; AAD13968.1; -.
DR   PIR; S01896; S01896.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120920; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Sushi; Alternative splicing.
FT   SIGNAL       1     34
FT   CHAIN       35    377       MEMBRANE COFACTOR PROTEIN.
FT   DOMAIN      35    328       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    329   351       POTENTIAL.
FT   DOMAIN      352   377       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       35    95       SUSHI 1.
FT   DOMAIN       98   158       SUSHI 2.
FT   DOMAIN      161   224       SUSHI 3.
FT   DOMAIN      227   284       SUSHI 4.
FT   DOMAIN      287   311       SER/THR-RICH.
FT   DISULFID     35    80       BY SIMILARITY.
FT   DISULFID     64    94       BY SIMILARITY.
FT   DISULFID     99   141       BY SIMILARITY.
FT   DISULFID    127   157       BY SIMILARITY.
FT   DISULFID    162   210       BY SIMILARITY.
FT   DISULFID    191   223       BY SIMILARITY.
FT   DISULFID    228   270       BY SIMILARITY.
FT   DISULFID    256   283       BY SIMILARITY.
FT   CARBOHYD     83    83       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    114   114       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    273   273       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    286   300       MISSING (IN A SECOND FORM).
FT   VARSPLIC    340   352       MISSING (IN ISOFORM M).
FT   VARSPLIC    353   361       YIQWRKKKQ -> DIEEKGGRKCKQMVELNMPLIRLNQPLQ
FT                               QSREAE (IN ISOFORM M).
FT   VARSPLIC    362   377       IYLIDEIHREVEFISL -> KAPGGAEYAIYQIKSIFPAEQ
FT                               EG (IN ISOFORM M AND ISOFORM N).
FT   VARSPLIC    362   377       MISSING (IN ISOFORM M AND ISOFORM N).
FT   VARSPLIC    340   361       MISSING (IN ISOFORM N).
FT   VARSPLIC    339   339       I -> IGKQMVELNMPLIRLNQPLQQSREAE (IN
FT                               ISOFORM N).
SQ   SEQUENCE   377 AA;  42247 MW;  2CA6F6125257DB57 CRC64;

  Query Match             25.1%;  Score 188.5;  DB 1;  Length 377;
  Best Local Similarity   30.6%;  Pred. No. 1.1e-11;
  Matches   41;  Conservative   24;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCPPPPEVENARPFPYYSLPIVPG TVLPYTCSPS-   YRLTGEKAIFCISENQVHA 54
           :: | |||:::|| :  :  : :       : |:| |:     : :||  |:|  :| |
Db     159 KVLCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYC GDNSV 215

Qy      55 IWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAIFRHGDSVTFICKANFTMKGSKIVWC 114
             | :| | |    | : |  |:|  |    |   | :  :|  :   | : :| | |: |
Db     216 -WSRAAPEC  --KVVKCPFPVVENGKQISGFGKEFYYRAIVMEE-EGGFYLIGSPTIVC 270

Qy     115 QANEMWGPIALPVC 128
             :.  | |  :| |
Db     271 DSNSIWDP PVPKC 283


RESULT 10
CFAH_MOUSE
ID   CFAH_MOUSE     STANDARD;      PRT;  1244 AA.
AC   P06909;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement factor H precursor (Protein beta-1-H).
GN   HF1 OR CFH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86233354; PubMed=2940596;
RA   Kristensen T., Tack B.F.;
RT   "Murine protein H is comprised of 20 repeating units, 61 amino acids
RT   in length.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN   [2]
```

```
RP   SEQUENCE OF 1-19 FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=90148935; PubMed=2533512;
RA   Munoz-Canoves P., Tack B.F., Vik D.P.;
RT   "Analysis of complement factor H mRNA expression: dexamethasone and
RT   IFN-gamma increase the level of H in L cells.";
RL   Biochemistry 28:9891-9897(1989).
RN   [4]
RP   SEQUENCE OF 1-18 FROM N.A.
RX   MEDLINE=90111033; PubMed=2136885;
RA   Natsuume-Sakai S., Nonaka M., Nonaka M., Harada Y.N., Shreffler D.C.,
RA   Moriwaki K.;
RT   "Demonstration of an unusual allelic variation of mouse factor H by
RT   the complete cDNA sequence of the H.2 allotype.";
RL   J. Immunol. 144:358-362(1990).
CC   -!- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC       C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC       C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC       CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC   -!- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
CC       MICE.
CC   -!- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch)
CC   ---------------------------------------------------------------------
DR   EMBL; M12660; AAA37759.1;  -.
DR   EMBL; J02891; AAA37795.1;  -.
DR   EMBL; M31979; AAA37762.1;  -.
DR   PIR; A26154; NBMSH.
DR   HSSP; P08603; 1HFI.
DR   MGD; MGI:88385; C1h.
DR   InterPro; IPR000436; Sushi_SCR_CCP
DR   Pfam; PF00084; sushi; 20.
DR   SMART; SM00032; CCP; 20.
KW   Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW   Signal.
FT   SIGNAL       1     18       BY SIMILARITY.
FT   CHAIN       19   1234       COMPLEMENT FACTOR H.
FT   DOMAIN      20     81       SUSHI 1.
FT   DOMAIN      84    142       SUSHI 2.
FT   DOMAIN     145    206       SUSHI 3.
FT   DOMAIN     209    263       SUSHI 4.
FT   DOMAIN     266    321       SUSHI 5.
FT   DOMAIN     324    386       SUSHI 6.
FT   DOMAIN     388    443       SUSHI 7.
FT   DOMAIN     447    506       SUSHI 8.
FT   DOMAIN     508    565       SUSHI 9.
FT   DOMAIN     568    623       SUSHI 10.
FT   DOMAIN     628    684       SUSHI 11.
FT   DOMAIN     689    744       SUSHI 12.
FT   DOMAIN     751    803       SUSHI 13.
FT   DOMAIN     807    862       SUSHI 14.
FT   DOMAIN     866    932       SUSHI 15.
FT   DOMAIN     935    990       SUSHI 16.
FT   DOMAIN     993   1049       SUSHI 17
FT   DOMAIN    1052   1108       SUSHI 18
FT   DOMAIN    1113   1169       SUSHI 19.
FT   DOMAIN    1171   1234       SUSHI 20.
FT   DISULFID    21     66       BY SIMILARITY.
FT   DISULFID    52     80       BY SIMILARITY.
FT   DISULFID    85    129       BY SIMILARITY.
FT   DISULFID   114    141       BY SIMILARITY.
FT   DISULFID   146    192       BY SIMILARITY.
FT   DISULFID   178    205       BY SIMILARITY.
FT   DISULFID   210    251       BY SIMILARITY.
FT   DISULFID   247    262       BY SIMILARITY.
FT   DISULFID   267    309       BY SIMILARITY.
FT   DISULFID   294    320       BY SIMILARITY.
FT   DISULFID   325    374       BY SIMILARITY.
FT   DISULFID   357    385       BY SIMILARITY.
FT   DISULFID   389    431       BY SIMILARITY.
FT   DISULFID   416    442       BY SIMILARITY.
FT   DISULFID   448    494       BY SIMILARITY.
FT   DISULFID   477    505       BY SIMILARITY.
FT   DISULFID   509    553       BY SIMILARITY.
FT   DISULFID   536    564       BY SIMILARITY.
FT   DISULFID   569    610       BY SIMILARITY.
FT   DISULFID   597    622       BY SIMILARITY.
FT   DISULFID   629    672       BY SIMILARITY.
FT   DISULFID   658    683       BY SIMILARITY.
FT   DISULFID   690    732       BY SIMILARITY.
FT   DISULFID   718    743       BY SIMILARITY.
FT   DISULFID   752    791       BY SIMILARITY.
FT   DISULFID   780    802       BY SIMILARITY.
FT   DISULFID   808    850       BY SIMILARITY.
FT   DISULFID   836    861       BY SIMILARITY.
FT   DISULFID   867    920       BY SIMILARITY.
FT   DISULFID   906    931       BY SIMILARITY.
FT   DISULFID   936    978       BY SIMILARITY.
FT   DISULFID   964    989       BY SIMILARITY.
FT   DISULFID   994   1037       BY SIMILARITY.
FT   DISULFID  1023   1048       BY SIMILARITY.
FT   DISULFID  1053   1096       BY SIMILARITY.
FT   DISULFID  1082   1107       BY SIMILARITY.
FT   DISULFID  1114   1157       BY SIMILARITY.
FT   DISULFID  1143   1168       BY SIMILARITY.
FT   DISULFID  1172   1223       BY SIMILARITY.
FT   DISULFID  1206   1233       BY SIMILARITY.
FT   CARBOHYD   676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   773    773       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   801    801       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1030   1030       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1225   1225       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1234 AA;  139082 MW;  C5AC02F341B957F7 CRC64;

  Query Match             23.4%;  Score 57;  DB 1;  Length 1234;
  Best Local Similarity   37.4%;  Pred No. 6.9e-10;
  Matches   52;  Conservative   15;  Mismatches   50;  Indels   22;  Gaps    9;

Qy        3 SCDPPEVKN-----ARKPYYSEPIVPGTVLEYTCSPSYRLIGEKAIFCISENQVHAIWD 57
            || || |: |     |:|  ||     | |  : | | : | |   | |:  |    |
Db      628 SCAPPLEILNGEINGAKKVEYS----HGEVVPVDCKPRFLLAGPNKIQCVDGN-    WT 678

Qy       58 KAPPICESVNKTISCSD-PIVPGGEMNKGSKAPFRHGDSVTFICKANFIMKGSKIVWCQA 116
             |:|    :|  | | | : |   | |   : || || | |: |||| |  :| | :
Db      679 TL-PVCIEEERT--CGDIPELEHG-SAKCSVIPYHHGDSVEFICEENFIMIGHGSVSCIS 734

Qy      117 NEMWGPTALPVCESDFPLE 135
              : |  | || | :   ||
Db      735 GK-W--TQLPKCVATDQLE 750


RESULT 11
CFAH_BOVIN
ID   CFAH_BOVIN     STANDARD;      PRT:  685 AA.
AC   Q28085;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement factor H (H factor 1) (Fragments).
GN   BF1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.
```

```
RC   TISSUE=Liver;
RX   MEDLINE=96202005; PubMed=8615824;
RA   Soames C.J., Day A.J., Sim R.B.;
RT   "Prediction from sequence comparisons of residues of factor H involved
RT   in the interaction with complement component C4b.";
RL   Biochem. J. 315:523-531(1996).
CC   -!- FUNCTION: Factor H functions as a cofactor in the inactivation of
CC       C4b by factor I and also increases the rate of dissociation of the
CC       C3bBb complex (C3 convertase) and the (C3b)NBB complex (C5
CC       convertase) in the alternative complement pathway (By similarity).
CC   -!- SIMILARITY: CONTAINS AT LEAST 13 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X98697; CAA67257.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 11.
DR   SMART; SM00032; CCP; 11.
KW   Complement alternate pathway; Plasma; Repeat; Sushi.
FT   NON_TER       1      1
FT   NON_CONS     16     17
FT   DOMAIN       17     67       SUSHI 2.
FT   DOMAIN       70    131       SUSHI 3.
FT   DOMAIN      134    188       SUSHI 4.
FT   DOMAIN      191    246       SUSHI 5.
FT   DOMAIN      249    309       SUSHI 6.
FT   DOMAIN      311    366       SUSHI 7.
FT   DOMAIN      370    429       SUSHI 8.
FT   DOMAIN      431    488       SUSHI 9.
FT   DOMAIN      491    547       SUSHI 10.
FT   DOMAIN      554    609       SUSHI 11.
FT   DOMAIN      614    668       SUSHI 12.
FT   DOMAIN      675    685       SUSHI 13.
FT   DISULFID     39     66       BY SIMILARITY.
FT   DISULFID     71    117       BY SIMILARITY.
FT   DISULFID    103    130       BY SIMILARITY.
FT   DISULFID    145    176       BY SIMILARITY.
FT   DISULFID    162    187       BY SIMILARITY.
FT   DISULFID    192    234       BY SIMILARITY.
FT   DISULFID    219    245       BY SIMILARITY.
FT   DISULFID    250    297       BY SIMILARITY.
FT   DISULFID    280    308       BY SIMILARITY.
FT   DISULFID    312    354       BY SIMILARITY.
FT   DISULFID    339    365       BY SIMILARITY.
FT   DISULFID    371    417       BY SIMILARITY.
FT   DISULFID    400    428       BY SIMILARITY.
FT   DISULFID    432    476       BY SIMILARITY.
FT   DISULFID    459    487       BY SIMILARITY.
FT   DISULFID    492    534       BY SIMILARITY.
FT   DISULFID    520    546       BY SIMILARITY.
FT   DISULFID    554    597       BY SIMILARITY.
FT   DISULFID    583    608       BY SIMILARITY.
FT   DISULFID    615    656       BY SIMILARITY.
FT   DISULFID    642    667       BY SIMILARITY.
FT   NON_TER     685    685
SQ   SEQUENCE   685 AA;  77506 MW;  69FC9DC8D530E872 CRC64;

  Query Match             20.9%;  Score 157.5;  DB 1;  Length 685;
  Best Local Similarity   31.3%;  Pred. No. 2.6e-08;
  Matches   42;  Conservative   14;  Mismatches   51;  Indels   27;  Gaps    6;

Qy        3 SCDPPPEVKNA RKPYYSLEYEVGTEVRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 61
            :|   ||:| | ||  :      |: | |:| : : |   | |:       | |
Db      553 SCALPPELPNGKRKETHKEEYAHNEVVEYACNEPEFLMKGSHKIQCVDGEST------ALW 606

Qy       62 ICESVNKTISCSD -      PIVPGGEMNKGSKAPFPHPGSVIFTCKANFIMKGSKIVWC 114
```

```
           :|    ::   ||      | ||          | :|||   : :   |:   : : |
Db      607 VCIEEERTCEISDLIPKDVPKPSVP    -      PIHKRDSVFISCKEAFIMIGPRTTC 656

Qy      115 QANEMWGPTALPVC 128
            : | |   |  | |
Db      657 ISGE W  TQPPQC 667


RESULT 12
C4BB_HUMAN
ID   C4BB_HUMAN     STANDARD;      PRT;   252 AA.
AC   P20851;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   C4b-binding protein beta chain precursor.
GN   C4BPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93415479; PubMed=8325877;
RA   Hillarp A., Pardo-Manuel F., Ruiz R., de Cordoba S., Dahlback B.;
RT   "The human C4b-binding protein beta chain gene.";
RL   J. Biol. Chem. 268:15017-15023(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90138962; PubMed=2300577;
RA   Hillarp A., Dahlback B.;
RT   "Cloning of cDNA coding for the beta chain of human complement
RT   component C4b-binding protein: sequence homology with the alpha
RT   chain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1183-1187(1990).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C4B/C4a INACTIVATOR
CC       (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. IT ALSO
CC       INTERACTS WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID P
CC       COMPONENT. THE BETA CHAIN BINDS PROTEIN S.
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC       OF 8 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1
CC       BETA CHAIN, A 540 kDa HOMOHEPTAMER OF ALPHA CHAINS OR A 500 kDa
CC       COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC       THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC       BINDING SITE FOR C4B AT THE END.
CC   -!- SIMILARITY: TO C4BP ALPHA CHAIN AND TO PIG APOLIPOPROTEIN H.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L11244; AAA35615.1; -.
DR   EMBL; L11245; AAA35616.1; -.
DR   EMBL; M29964; AAB59520.1; -.
DR   PIR; A34877; A34877.
DR   PIR; A47107; A47107.
DR   HSSP; P10998; 1VVD.
DR   MIM; 120831; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW   Polymorphism.
FT   SIGNAL        1     17
FT   CHAIN        18    252       C4B-BINDING PROTEIN BETA CHAIN
```

```
FT   DOMAIN       22     77       SUSHI 1.
FT   DOMAIN       80    135       SUSHI 2.
FT   DOMAIN      138    192       SUSHI 4.
FT   DISULFID     23     63       BY SIMILARITY.
FT   DISULFID     49     76       BY SIMILARITY.
FT   DISULFID     81    121       BY SIMILARITY.
FT   DISULFID    107    134       BY SIMILARITY.
FT   DISULFID    139    179       BY SIMILARITY.
FT   DISULFID    165    191       BY SIMILARITY.
FT   DISULFID    202    202       INTERCHAIN (WITH ALPHA CHAIN)
FT                                (POTENTIAL).
FT   DISULFID    216    216       INTERCHAIN (WITH ALPHA CHAIN)
FT                                (POTENTIAL).
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    117    117       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     198    198       P -> S (IN DBSNP:1803226).
FT                                /FTId=VAR_012039.
SQ   SEQUENCE   252 AA;  28357 MW;  0F6CC64067C5E2E7 CRC64;

  Query Match             20.8%;  Score 156.5;  DB 1;  Length 252;
  Best Local Similarity   28.9%;  Pred. No. 1.1e-08;
  Matches   47;  Conservative   19;  Mismatches   57;  Indels   15;  Gaps    6;

Qy       4 CDPPPEVKNARKPYYSLPIVPGTVL RYTCSPSYRLIGEKAIFCISENQVHATWDKAPPI 62
           |   | | |:     :   | | :| | |   | |:|:| :|| :   :    ||
Db      28 CPELPPVDNS---IFVAKEVEGQILGTYVCIKGYHLVGKKTLFCNASKE---WDNTTTE 75

Qy      63 CESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWGP 122
           |    :   | ||::  |  :  | |    | :|| |  :: :|||    |  :  | |
Db      76 C-  --RLGHCPDPVLVNGEFS--SSGPVNVSDKITFMCNDHYILKGSNRSQCLEDHTWAP 129

Qy     123 TALPVCES 130
             |:|:|
Db     140 -PEPICKS 136


RESULT 13
DAF2_MOUSE
ID   DAF2_MOUSE     STANDARD;      PRT;   407 AA.
AC   Q61476;
DT   01-NOV 1997 (Rel. 35, Created)
DT   01-NOV 1997 (Rel. 35, Last sequence update)
DT   01-NOV 1997 (Rel. 35, Last annotation update)
DE   Complement decay-accelerating factor, transmembrane precursor
DE   (DAF TM).
GN   DAF2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=95404982; PubMed=7545711;
RA   Spicer A.P., Seldin M.F., Gendler S.J.;
RT   "Molecular cloning and chromosomal localization of the mouse decay-
RT   accelerating factor genes. Duplicated genes encode
RT   glycosylphosphatidylinositol anchored and transmembrane forms.";
RL   J. Immunol. 155:3079-3091(1995).
CC   -!- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: TESTES, SPLEEN AND LYMPH NODE.
CC   -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------- ---- ---------- ----------------------------------------

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---- ---------- ----------------------------------------
DR   EMBL; L41365; AAB00092.1; -.
DR   HSSP; P08603; 1HCC.
DR   MGD; MGI:104849; Daf2.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Repeat; Signal; Sushi;
KW   Transmembrane
FT   SIGNAL        1     39       POTENTIAL.
FT   CHAIN        40    407       COMPLEMENT DECAY-ACCELERATING FACTOR,
FT                                TRANSMEMBRANE.
FT   DOMAIN        40    368       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     369    389       POTENTIAL
FT   DOMAIN       390    407       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        40    100       SUSHI 1.
FT   DOMAIN       102    164       SUSHI 2.
FT   DOMAIN       167    226       SUSHI 3.
FT   DOMAIN       229    290       SUSHI 4.
FT   DOMAIN       291    363       SER/THR RICH (BY SIMILARITY).
FT   DISULFID      70     99       BY SIMILARITY.
FT   DISULFID     103    150       BY SIMILARITY.
FT   DISULFID     134    163       BY SIMILARITY.
FT   DISULFID     158    209       BY SIMILARITY.
FT   DISULFID     195    225       BY SIMILARITY.
FT   DISULFID     230    272       BY SIMILARITY.
FT   DISULFID     258    289       BY SIMILARITY.
FT   CARBOHYD     132    192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   407 AA;  44469 MW;  DDD82F2D72CEAD40 CRC64;

  Query Match             20.4%;  Score 153.5;  DB 1;  Length 407;
  Best Local Similarity   27.5%;  Pred. No. 3.7e-08;
  Matches   39;  Conservative   31;  Mismatches   53;  Indels   19;  Gaps    8;

Qy       3 SCDPPPEVKNARKPYYSLP--IVPGTVLRYTCS PSYRLIGEKAIFC-ISENQVHAIWDKA 59
           ||  | :: |    : ::|  |: |: : ::| | |||:|  :| | |: | |   ||
Db     167 SCPNPKDLDNG---HINIPTGILFGSEINFSCN PGYRLVGITSILCTITGNAV--DWDDE 221

Qy      60 PPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQANE 118
             |:|       | | ||      : :|    ::    | : |   | : |: |::|  ::
Db     222 FPVCTE----IFCPDPPKINDGIMRGESDSYKYGQVVTFSCDKGFILEGNSIIYCIVSKS 277

Qy     119 ---MWGPTALPVC--ESDFPLE 135
               |  :  | |  ||  |::
Db     278 DVGQWS-SPPPQCIEESKVPIK 298


RESULT 14
CFAB_HUMAN
ID   CFAB_HUMAN     STANDARD;      PRT;   764 AA.
AC   P00751; O15006; Q29944; Q9BTF5; Q9BX92;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B precursor (EC 3.4.21.47) (C3/C5 convertase)
DE   (Properdin factor B) (Glycine-rich beta glycoprotein) (GBG) (PBF2).
GN   BF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ALLELES S; FA AND FB).
RX   MEDLINE=91065702; PubMed=2249879;
RA   Davrinche C., Abbal M., Clerc A.;
```

```
RT   "Molecular characterization of human complement factor B subtypes.";
RL   Immunogenetics 32:409 412(1990).
RN   [2]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=Liver;
RX   MEDLINE=94237745; PubMed=8186962;
RA   Mejia J.E., Jahn I., de la Salle H., Hauptmann G.;
RT   "Human factor B: Complete cDNA sequence of the BF*S allele.";
RL   Hum. Immunol. 39:49 53(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ALLELE S).
RC   TISSUE=Liver;
RX   MEDLINE=94041899; PubMed=8225086;
RA   Schwaeble W., Luettig B., Sokolowski T., Estaller C., Weiss E.H.,
RA   Meyer Zum Buschenfelde K.-H., Whaley K., Dippold W.;
RT   "Human complement factor B: functional properties of a recombinant
RT   zymogen of the alternative activation pathway convertase.";
RL   Immunobiology 188:221 232(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ALLELE S).
RX   MEDLINE=94006177; PubMed=8243029;
RA   Bortuchi I., Kim S., Matsumoto M., Watanabe I., Fujita S.,
RA   Volanakis J.E.;
RT   "Human complement factor B: cDNA cloning, nucleotide sequencing,
RT   phenotypic conversion by site directed mutagenesis and expression.";
RL   Mol. Immunol. 30:1587 1592(1993).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Bowen T., Oankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT   "Sequence determination of 400 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (SEP 1997) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (MAR 2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Jaatinen T., Kaneva J., Poutanen K.E., Saarinen Pihkala U.,
RA   Lokki M.-L.;
RT   "Expression and alternative splicing of human factor B gene in
RT   leukemic mononuclear cells.";
RL   Submitted (FEB 2001) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 26 764, PARTIAL SEQUENCE FROM N.A., AND CARBOHYDRATES.
RX   MEDLINE=84161997; PubMed=6546754;
RA   Mole J.E., Anderson J.K., Davison E.A., Woods D.E.;
RT   "Complete primary structure for the zymogen of human complement
RT   factor B.";
RL   J. Biol. Chem. 259:3407 3412(1984).
RN   [9]
RP   SEQUENCE OF 260 764.
RX   MEDLINE=83204002; PubMed=6342610;
RA   Christie D.L., Gagnon J.;
RT   "Amino acid sequence of the Bb fragment from complement factor B.
RT   Sequence of the major cyanogen bromide cleavage peptide (CB-11) and
RT   completion of the sequence of the Bb fragment.";
RL   Biochem. J. 209:61 70(1983).
RN   [10]
RP   SEQUENCE OF 339 764 FROM N.A.
RX   MEDLINE=83273641; PubMed=6308626;
RA   Campbell R.D., Porter R.R.;
RT   "Molecular cloning and characterization of the gene coding for human
RT   complement protein factor B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 80:4464 4468(1983).
RN   [11]
RP   SEQUENCE OF 467 595 AND 75 764 FROM N.A.
RX   MEDLINE=84039428; PubMed=6092884;
RA   Woods D.E., Markham A.F., Ricker A.T., Goldberger G., Colten H.R.;
RT   "Isolation of cDNA clones for the human complement protein factor B,
RT   a class III major histocompatibility complex gene product.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:5661 5665(1982).
RN   [12]
RP   SEQUENCE OF 16-259 FROM N.A.
RX   MEDLINE=84158524; PubMed=6323161;
RA   Morley B.J., Campbell R.D.;
RT   "Internal homologies of the Ba fragment from human complement
RT   component Factor B, a class III MHC antigen.";
RL   EMBO J. 3:153 157(1984).
RN   [13]
RP   SEQUENCE OF 1-99 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=87102880; PubMed=3643061;
RA   Wu L.C., Morley B.J., Campbell R.D.;
RT   "Cell-specific expression of the human complement protein factor B
RT   gene: evidence for the role of two distinct 5' flanking elements.";
RL   Cell 48:331-342(1987).
RN   [14]
RP   GLYCATION IN POSITION 291.
RX   MEDLINE=91174758; PubMed=2006911;
RA   Niemann M.A., Bhown A.S., Miller E.J.;
RT   "The principal site of glycation of human complement factor B.";
RL   Biochem. J. 274:473-480(1991).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE. IT HAS ALSO BEEN IMPLICATED
CC       IN PROLIFERATION AND DIFFERENTIATION OF PREACTIVATED B
CC       LYMPHOCYTES, RAPID SPREADING OF PERIPHERAL BLOOD MONOCYTES,
CC       STIMULATION OF LYMPHOCYTE BLASTOGENESIS AND LYSIS OF ERYTHROCYTES.
CC       BA INHIBITS THE PROLIFERATION OF PREACTIVATED B LYMPHOCYTES.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha chain to yield C5a and C5b. Both
CC       cleavages take place at the C-terminal of an arginine residue.
CC   -!- SUBUNIT: MONOMER.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- POLYMORPHISM: TWO MAJOR VARIANTS, F AND S, AND 2 MINOR VARIANTS,
CC       AS WELL AS AT LEAST 14 VERY RARE VARIANTS, HAVE BEEN IDENTIFIED.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------- . . . . . .
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- . . . . . .
DR   EMBL; X72875; CAA51489.1; -.
DR   EMBL; S67410; AAD13989.1; -.
DR   EMBL; L15702; AAA16820.1; -.
DR   EMBL; X00284; CAA25077.1; -.
DR   EMBL; AF019413; AAB67977.1; -.
DR   EMBL; BC004143; AAH04143.1; -.
DR   EMBL; AF349679; AAK40167.1; -.
DR   EMBL; K01566; AAA36225.2; -.
DR   EMBL; J00125; -; NOT_ANNOTATED_CDS.
DR   EMBL; J00126; AAA36226.1; -.
DR   EMBL; J00185; AAA36219.1; ALT_SEQ.
DR   EMBL; J00186; AAA36220.1; -.
DR   EMBL; M15082; AAA59625.1; -.
DR   PIR; A00944; BBHU.
DR   PIR; S14339; S14339.
DR   PIR; S34075; S34075.
DR   HSSP; P00734; 2HNT.
DR   MEROPS; S01.196; -.
DR   SWISS-2DPAGE; P00751; HUMAN.
```

```
DR   Siena 2DPAGE; P00751;  .
DR   MIM; 138470;  .
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 3.
DR   Pfam; PF00089; trypsin; 1.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi; Signal; Polymorphism; Zymogen;
KW   Alternative splicing.
FT   SIGNAL       1     25
FT   CHAIN       26    764       COMPLEMENT FACTOR B.
FT   CHAIN       26    259       BA FRAGMENT.
FT   CHAIN      260    764       BB FRAGMENT.
FT   DOMAIN      36     99       SUSHI 1.
FT   DOMAIN     102    159       SUSHI 2.
FT   DOMAIN     164    219       SUSHI 3.
FT   DOMAIN     270    469       VWFA.
FT   DOMAIN     482    764       SERINE PROTEASE.
FT   ACT_SITE   526    526       CHARGE RELAY SYSTEM.
FT   ACT_SITE   576    576       CHARGE RELAY SYSTEM.
FT   ACT_SITE   699    699       CHARGE RELAY SYSTEM.
FT   DISULFID    37     76       BY SIMILARITY.
FT   DISULFID    62     98       BY SIMILARITY.
FT   DISULFID   103    145       BY SIMILARITY.
FT   DISULFID   131    158       BY SIMILARITY
FT   DISULFID   165    205       BY SIMILARITY.

  Query Match             20.4%;  Score 153.5;  DB 1;  Length 764;
  Best Local Similarity   31.1%;  Pred. No. 7.4e-08;
  Matches   42;  Conservative   17;  Mismatches   59;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA----RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           | |  | : :|     | |||::       : : |   | | |     |    ||:  |
Db     101 IHCPRPHDFENGEYWPRSPYYNV--SDEISFHCYDGYTLRGSANRTC----QVNGRWS 152

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQAN 117
             ||:  |   ||:| :| |    ||:  :|  ||||: |    |::||:   ||
Db     153 GQTAICD--NGAGYCSNPGIPIGTRKVGSQ--YRLEDSVTYHCSRGLTLRGSQRRTCQEG 208

Qy     118 EMWGPTALPVCESDF 132
           |  |  | |:  |
Db     209 GSWSGTE-PSCQDSF 222


RESULT 15
CFAH_HUMAN
ID   CFAH_HUMAN     STANDARD;      PRT;  1231 AA.
AC   P08603; Q14570; P78445; Q9NU86;
DT   01 AUG 1988 (Rel. 08, Created)
DT   01 JAN 1990 (Rel. 13, Last sequence update)
DT   16-OCT 2001 (Rel. 40, Last annotation update)
DE   Complement factor H precursor (H factor 1).
GN   HF1 OR HF OR CFH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.
RC   TISSUE=Liver;
RX   MEDLINE=88134059; PubMed=2963625;
RA   Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;
RT   "The complete amino acid sequence of human complement factor H.";
RL   Biochem. J. 249:593-602(1988).
RN   [2]
RP   SEQUENCE OF 53-445 FROM N.A.
RX   MEDLINE=87054207; PubMed=2946589;
RA   Schulz T.F., Schwaeble W., Stanley K.K., Weiss E., Dierich M.P.;
RT   "Human complement factor H: isolation of cDNA clones and partial cDNA
RT   sequence of the 38-kDa tryptic fragment containing the binding site
RT   for C3b.";
RL   Eur. J. Immunol. 16:1351-1355(1986).
RN   [3]
RP   SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=86169701; PubMed=2937845;
RA   Kristensen T., Wetsel R.A., Tack B.F.;
RT   "Structural analysis of human complement protein H: homology with C4b
RT   binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";
RL   J. Immunol. 136:3407-3411(1986).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Bird C.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1047-1231 FROM N.A.
RX   MEDLINE=91201892; PubMed=1826708;
RA   Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;
RT   "Cloning of the 1.4-kb mRNA species of human complement factor H
RT   reveals a novel member of the short consensus repeat family related
RT   to the carboxy terminal of the classical 150 kDa molecule.";
RL   J. Immunol. 146:3190-3196(1991).
RN   [6]
RP   SEQUENCE OF 19-35.
RX   MEDLINE=83048213; PubMed=6215918;
RA   Sim R.B., Discipio R.G.;
RT   "Purification and structural studies of the complement system control
RT   protein beta 1H (Factor H).";
RL   Biochem. J. 205:285-293(1982).
RN   [7]
RP   SEQUENCE OF 1-19 FROM N.A.
RA   Vik D.P., Williams S.A.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 1-9 FROM N.A.
RA   Dominquez O.;
RL   Thesis (1993), Hospital Trias I Pujol, Spain.
RN   [9]
RP   STRUCTURE BY NMR OF 927-985 (SUSHI 16).
RX   MEDLINE=91278097; PubMed=1829116;
RA   Norman D.G., Barlow P.N., Baron M., Day A.J., Sim R.B., Campbell I.D.;
RT   "Three-dimensional structure of a complement control protein module
RT   in solution.";
RL   J. Mol. Biol. 219:717-725(1991).
RN   [10]
RP   STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX   MEDLINE=92242549; PubMed=1533152;
RA   Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
RA   Driscoll P.C., Sim R.B., Campbell I.D.;
RT   "Solution structure of the fifth repeat of factor H: a second example
RT   of the complement control protein module.";
RL   Biochemistry 31:3626-3634(1992).
RN   [11]
RP   STRUCTURE BY NMR OF 866-985 (SUSHIS 15 AND 16).
RX   MEDLINE=93323119; PubMed=8331663;
RA   Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RA   Sim R.B., Campbell I.D.;
RT   "Solution structure of a pair of complement modules by nuclear
RT   magnetic resonance.";
RL   J. Mol. Biol. 232:268-284(1993).
CC   -!- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC       C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC       C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3B)NBB COMPLEX (C5
CC       CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
```

```
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC     -!- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.
CC     -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC         441 ONWARD DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00716; CAA68704.1;
DR   EMBL; X04697; CAB41739.1; ALT_FRAME.
DR   EMBL; X07523; CAA30403.1; -.
DR   EMBL; M12883; AAA52018.1; -.
DR   EMBL; AL049744; CAB70598.1; -.
DR   EMBL; M65294; AAA35948.1; -.
DR   EMBL; U56979; AAB01987.1; -.
DR   EMBL; Z29665; CAA82764.1; -.
DR   PIR; S00254; NIRHUE.
DR   PIR; S04013; S04013.
DR   PDB; 1HCF; 15-APR-92.
DR   PDB; 1HFH; 15-JUL-93.
DR   PDB; 1HFI; 15-JUL-93.
DR   MIM; 134370; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 20.
DR   SMART; SM00032; CCP; 20.
KW   Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW   Signal; 3D-structure; Polymorphism; Alternative splicing.
FT   SIGNAL       1     18
FT   CHAIN       19   1231       COMPLEMENT FACTOR H.
FT   DOMAIN      20     81       SUSHI 1.
FT   DOMAIN      84    142       SUSHI 2.
FT   DOMAIN     145    206       SUSHI 3.
FT   DOMAIN     209    263       SUSHI 4.
FT   DOMAIN     266    321       SUSHI 5.
FT   DOMAIN     324    386       SUSHI 6.
FT   DOMAIN     388    444       SUSHI 7.
FT   DOMAIN     447    506       SUSHI 8.
FT   DOMAIN     508    565       SUSHI 9.
FT   DOMAIN     568    624       SUSHI 10.
FT   DOMAIN     629    685       SUSHI 11.
FT   DOMAIN     690    745       SUSHI 12.
FT   DOMAIN     752    804       SUSHI 13.
FT   DOMAIN     810    865       SUSHI 14.
FT   DOMAIN     869    927       SUSHI 15.
FT   DOMAIN     930    985       SUSHI 16.
FT   DOMAIN     988   1044       SUSHI 17.
FT   DOMAIN    1047   1103       SUSHI 18.
FT   DOMAIN    1108   1164       SUSHI 19.
FT   DOMAIN    1166   1229       SUSHI 20.
FT   DISULFID    21     66       BY SIMILARITY.
FT   DISULFID    52     80       BY SIMILARITY.
FT   DISULFID    85    129       BY SIMILARITY.
FT   DISULFID   114    141       BY SIMILARITY.
FT   DISULFID   146    192       BY SIMILARITY.
FT   DISULFID   178    205       BY SIMILARITY.
FT   DISULFID   210    251       BY SIMILARITY.
FT   DISULFID   237    262       BY SIMILARITY.
FT   DISULFID   267    309       BY SIMILARITY.
FT   DISULFID   294    320       BY SIMILARITY.
FT   DISULFID   325    374       BY SIMILARITY.
FT   DISULFID   357    385       BY SIMILARITY.
FT   DISULFID   389    431       BY SIMILARITY.
FT   DISULFID   416    442       BY SIMILARITY.
FT   DISULFID   448    494       BY SIMILARITY.
FT   DISULFID   477    505       BY SIMILARITY.
FT   DISULFID   509    554       BY SIMILARITY.
FT   DISULFID   536    564       BY SIMILARITY.
FT   DISULFID   569    611       BY SIMILARITY.
FT   DISULFID   597    623       BY SIMILARITY.
FT   DISULFID   640    673       BY SIMILARITY.
FT   DISULFID   659    684       BY SIMILARITY.
FT   DISULFID   691    733       BY SIMILARITY.
FT   DISULFID   719    744       BY SIMILARITY.
FT   DISULFID   753    792       BY SIMILARITY.
FT   DISULFID   781    803       BY SIMILARITY.
FT   DISULFID   811    853       BY SIMILARITY.
FT   DISULFID   839    864       BY SIMILARITY.
FT   DISULFID   870    915       BY SIMILARITY.
FT   DISULFID   901    926       BY SIMILARITY.
FT   DISULFID   941    973       BY SIMILARITY.
FT   DISULFID   959    984       BY SIMILARITY.
FT   DISULFID   989   1032       BY SIMILARITY.
FT   DISULFID  1018   1043       BY SIMILARITY.
FT   DISULFID  1048   1091       BY SIMILARITY.
FT   DISULFID  1077   1102       BY SIMILARITY.
FT   DISULFID  1109   1152       BY SIMILARITY.
FT   DISULFID  1148   1163       BY SIMILARITY.
FT   DISULFID  1167   1218       BY SIMILARITY.
FT   DISULFID  1201   1228       BY SIMILARITY.
FT   CARBOHYD   529    529       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   718    718       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   802    802       N-LINKED (GLCNAC...).
FT   CARBOHYD   822    822       N-LINKED (GLCNAC...).
FT   CARBOHYD   882    882       N-LINKED (GLCNAC...).
FT   CARBOHYD   911    911       N-LINKED (GLCNAC...).
FT   CARBOHYD  1029   1029       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD  1095   1095       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   446    449       KEPS -> SEEL (IN ISOFORM 2).
FT   VARSPLIC   450   1231       MISSING (IN ISOFORM 2).
FT   VARIANT    402    402       H -> Y.
FT                               /FTId=VAR_001979.
FT   CONFLICT    21     21       C -> Q (IN REF. 4).
FT   CONFLICT    30     30       I   -> V (IN REF. 4).
FT   CONFLICT    34     34       I   -> Q (IN REF. 4).
FT   CONFLICT    53     54       RD -> LL (IN REF. 2).
FT   STRAND     870    871
FT   STRAND     876    876
```

Query Match 19.8%; Score 149; DB 1; Length 1231;
Best Local Similarity 32.6%; Prod. No. 3.5e-07;
Matches 43; Conservative 18; Mismatches 49; Indels 22; Gaps 9;

```
Qy        3 SCDPPPE---- VKNARKPYYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
            || |||     ||   | |         |: | |:| : : |   | |     |  |
Db      629 SCGPPPEIINGNVKEKIKEFYQ----GSEVVEYYCNPRFLMKGPNKIQC--   VQGEWT 679

Qy       58 KAPPICESVNKCISCSD PIVPGGFMNKGSKAPFKKCDSVTFTCKANFTMKGSKIVWCQA 116
              :|  : :  :| | | :  |:    |  |: :|||| | |  :|:| | ::: |
Db      680 TL PVC--IVEESTCGDIPELEHGWAQL SSPPYYYGDSVEFNCSESFTMIGHRSITC I 734

Qy      117 NEMWGPTALPVC 128
            : :|  | |:|:
Db      735 HGVW--IQLPQC 744
```

RESULT 16
LEM2_RABIT
ID LEM2_RABIT STANDARD; PRT; 551 AA.
AC P27113;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=92189729; PubMed=1372169;
RA   Lariqan J.D., Tsang T.C., Rumberger J.M., Burns D.K.;
RT   "Characterization of cDNA and genomic sequences encoding rabbit
RT   ELAM 1: conservation of structure and functional interactions with
RT   leukocytes.";
RL   DNA Cell Biol. 11:149-162(1992).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM 1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS)
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- INDUCTION: BY CYTOKINES.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   ---------- . . ------ ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ---------- ----------
DR   EMBL; M91004; AAA31243.1; -.
DR   EMBL; M91005; AAA31244.1; -.
DR   HSSP; P16581; 1KJA.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL       1     24
FT   CHAIN       24    551       E-SELECTIN.
FT   DOMAIN      24    495       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    496   517       POTENTIAL.
FT   DOMAIN      518   551       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40   140       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      141   177       EGF-LIKE.
FT   DOMAIN      181   240       SUSHI 1.
FT   DOMAIN      243   302       SUSHI 2.
FT   DOMAIN      305   365       SUSHI 3.
FT   DOMAIN      368   428       SUSHI 4.
FT   DOMAIN      431   487       SUSHI 5.
FT   DISULFID     42   140       BY SIMILARITY.
FT   DISULFID    113   132       BY SIMILARITY.
FT   DISULFID    145   156       BY SIMILARITY.
FT   DISULFID    150   165       BY SIMILARITY.
FT   DISULFID    167   176       BY SIMILARITY.
FT   DISULFID    182   226       BY SIMILARITY.
FT   DISULFID    212   239       BY SIMILARITY.
FT   DISULFID    244   288       BY SIMILARITY.
FT   DISULFID    274   301       BY SIMILARITY.
FT   DISULFID    306   351       BY SIMILARITY.
FT   DISULFID    337   364       BY SIMILARITY.
FT   DISULFID    369   414       BY SIMILARITY.
FT   DISULFID    400   427       BY SIMILARITY.
FT   DISULFID    432   473       BY SIMILARITY.
FT   DISULFID    459   486       BY SIMILARITY.
FT   CARBOHYD     32    32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     45    45       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    162   162       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    194   194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201   201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    205   205       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267   267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    314   314       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321   321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    334   334       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    442   442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466   466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    308   308       T -> A (IN REF. 1; AAA31244).
FT   CONFLICT    328   328       T -> A (IN REF. 1; AAA31244).
FT   CONFLICT    491   491       A -> V (IN REF. 1; AAA31244).
SQ   SEQUENCE   551 AA;  60346 MW;  23B78A333B23240E CRC64;

  Query Match             19.7%;  Score 143.5;  DB 1;  Length 551;
  Best Local Similarity   25.9%;  Pred. No. 1.6e-07;
  Matches   36;  Conservative   20;  Mismatches   48;  Indels   45;  Gaps    6;

Qy       1 EISCDPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 60
           :: | |              |:          |   : |   : |:  ::: | |      :||
Db     254 DVKCSP----SQGSAPW-------NTTCTIDGEEGFILDGARSLQCTSS----GSWDNEK 298

Qy      61 PICESV---------NKTISCSDPIVPGGFMNKGSKAPFRHGDSVIFTCKANFTMKGSKI 111
           | |:::          | ::|||:           |:  |    |  |||: || :::|
Db     299 PTCKAVSCDTIHHPQNGSVSCSN----------SSEGKEIFRSSCNFTCEENFLLRGPAQ 348

Qy     112 VWCQANEMWGPTALPVCES 130
           | | |   |   | ||||:
Db     349 VECTAQGQWTQQA-PVCEA 366


RESULT 17
DAF1_MOUSE
ID   DAF1_MOUSE     STANDARD;      PRT;   390 AA.
AC   Q61475, Q61367, P97732,
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 38, Last annotation update)
DE   Complement decay-accelerating factor, GPI-anchored precursor
DE   (DAF-GPI).
GN   DAF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=95402982; PubMed=7545711;
RA   Spicer A.P., Seldin M.F., Gendler S.J.;
RT   "Molecular cloning and chromosomal localization of the mouse decay
RT   accelerating factor genes. Duplicated genes encode
RT   glycosylphosphatidylinositol-anchored and transmembrane forms.";
RL   J. Immunol. 155:3079-3091(1995).
RN   [2]
RP   SEQUENCE OF 7-390 FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Spleen;
RX   MEDLINE=96362213; PubMed=8671624;
RA   Fukuoka Y., Yasui A., Okada N., Okada H.;
RT   "Molecular cloning of murine decay accelerating factor by
RT   immunoscreening.";
RL   Int. Immunol. 8:379-385(1996).
CC   -!- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,
CC       LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.
CC   -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
```

```
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC    !  SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC    !  SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L41366; AAB00091.1; -.
DR   EMBL; D63679; BAA09830.1; -.
DR   HSSP; P08603; 1HCC.
DR   MGD; MGI:104850; Daf1.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Glycoprotein; Membrane; Repeat; GPI anchor;
KW   Signal; Sushi.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    361       COMPLEMENT DECAY ACCELERATING FACTOR,
FT                                GPI ANCHORED.
FT   PROPEP      362    390       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       35     95       SUSHI 1.
FT   DOMAIN       97    159       SUSHI 2.
FT   DOMAIN      162    221       SUSHI 3.
FT   DOMAIN      224    285       SUSHI 4.
FT   DOMAIN      288    364       SER/THR RICH.
FT   DISULFID     65     94       BY SIMILARITY
FT   DISULFID     98    145       BY SIMILARITY.
FT   DISULFID    129    158       BY SIMILARITY.
FT   DISULFID    163    204       BY SIMILARITY.
FT   DISULFID    190    220       BY SIMILARITY.
FT   DISULFID    225    267       BY SIMILARITY.
FT   DISULFID    258    284       BY SIMILARITY.
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       361    361       GPI-ANCHOR (BY SIMILARITY).
FT   CONFLICT      7      7       P  -> A (IN REF. 2).
FT   CONFLICT      9      9       I  -> A (IN REF. 2).
FT   CONFLICT     83     83       E  -> G (IN REF. 2).
FT   CONFLICT     91     91       E  -> G (IN REF. 2).
FT   CONFLICT    135    135       E  -> K (IN REF. 2).
FT   CONFLICT    173    173       H  -> L (IN REF. 2).
FT   CONFLICT    180    180       I  -> T (IN REF. 2).
SQ   SEQUENCE   390 AA;  42618 MW;  4418721DFF47E8E7 CRC64;

  Query Match             19.7%;  Score 148;  DB 1;  Length 390;
  Best Local Similarity   25.2%;  Pred. No. 1.3e-07;
  Matches   80;  Conservative   32;  Mismatches   45;  Indels   12;  Gaps    5;

Qy       4 SCDPPPEVKNARKPYYSLP  IVPGTVLRYTCSPSYRLIGEKAIFC ISENQVHAIWDKA 59
           ||    : |    :  :|  | : : : :: :| |||:|   : || :: | |    ||
Db     164 SCPNPKILDNG   RIITPTGTLEGSEINFSCNPGYRLVGVSSIFCSVTGNTV  DWDDE 216

Qy      60 PPICESVNKCISCSDPTVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQANE 118
             |:|  ::    | :|     : :|    : :   ||:||   | : |: :::|  ::
Db     217 EPVCAEIB    CPEPPKINNGIMKGDEDEYTYRSQVVTFYSCDKGFILVGNASIYCTVSK 271


RESULT 18
2AG_BIG
ID   C2AB_BIG       STANDARD;      PRT;   151 AA.
AC   Q03710;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B (EC 3.4.21.47) (C3/C5 convertase) (Properdin
DE   factor B) (Fragment).
GN   BF.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91372866; PubMed=1680099;
RA   Peelman L.J., van de Weghe A.P., Coppieters W.P., van Zeveren A.J.;
RA   Bouquet Y.H.;
RT   "Cloning and sequencing of the porcine complement factor B.";
RL   Immunogenetics 34:332-335(1991).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha-chain to yield C5a and C5b. Both
CC       cleavages take place at the C-terminal of an arginine residue.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS-III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M59240; AAA31021.1; -.
DR   HSSP; P02749; 1QUB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
DR   PROSITE; PS50240; TRYPSIN_DOM; PARTIAL.
DR   PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
DR   PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi.
FT   NON_TER       1      1
FT   DOMAIN        1     58       SUSHI 2.
FT   DOMAIN       63    118       SUSHI 3.
FT   DISULFID      4     46       BY SIMILARITY.
FT   DISULFID     32     59       BY SIMILARITY.
FT   DISULFID     66    106       BY SIMILARITY.
FT   DISULFID     92    119       BY SIMILARITY.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     151    151
SQ   SEQUENCE   151 AA;  16765 MW;  BD247E75047E517F CRC64;

  Query Match             19.2%;  Score 144.5;  DB 1;  Length 151;
  Best Local Similarity   30.4%;  Pred. No. 1e-07;
  Matches   41;  Conservative   17;  Mismatches   60;  Indels   17;  Gaps    6;

Qy       2 ISCPPPEVKNA----RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWD 57
           | |  | : :|    | |||:|      : : |  | | |       |  :|   ||
Db       2 IRCPKPHDFENGEYWPRAPYYNL----SDEISFHCYDGYTLRGSANRT    QVIGRWD 50

Qy      58 KAPLPICESVNKCISCSDPTVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKIVWCQAN 117
             ||:  :    | :| :| |    |::  :|  |||: |    |:|||:   ||
Db      54 GQTAICD--DGAGYCPNPGIPIGTRKVGTQ  YRLEDSVTYYCTRGLTLRGSQRPTCQEG 109

Qy     118 EMWGPTALPVCESDF 132
           |  :  | |:  |
Db     110 GSWSSTE-PSCQDSF 123
```

```
RESULT 19
LEM3_MOUSE
ID   LEM3_MOUSE      STANDARD;      PRT:   768 AA.
AC   Q01102;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP OR GRMP.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340571; PubMed=1378846;
RA   Weller A., Isenmann S., Vestweber D.;
RT   "Cloning of the mouse endothelial selectins. Expression of both E-
RT   and P-selectin is inducible by tumor necrosis factor alpha.";
RL   J. Biol. Chem. 267:15176-15183(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92345617; PubMed=1379089;
RA   Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT   "Molecular cloning and analysis of in vivo expression of murine P-
RT   selectin.";
RL   Blood 80:795-800(1992).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC       AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC       ACTIVATION BY AGONISTS, P SELECTIN IS TRANSPORTED RAPIDLY TO
CC       THE CELL SURFACE.
CC   -!- INDUCTION: BY TNF-ALPHA.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS; MOUSE P-LECTIN LACKS
CC       THE HUMAN SUSHI-2 EQUIVALENT.
CC   ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ----------
DR   EMBL; M87861; AAA40008.1; -.
DR   EMBL; M72382; AAA37712.1; -.
DR   PIR; A42755; A42755.
DR   HSSP; P16109; 1FSB.
DR   MGD; MGI:98280; Selp.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
FT   SIGNAL        1     41       POTENTIAL.
FT   CHAIN        42    768       P-SELECTIN.
FT   DOMAIN       42    709       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    710    733       POTENTIAL.
FT   DOMAIN      734    768       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      159    195       EGF-LIKE.
FT   DOMAIN      199    258       SUSHI 1.
FT   DOMAIN      261    320       SUSHI 2.
FT   DOMAIN      323    382       SUSHI 3.
FT   DOMAIN      385    444       SUSHI 4.
FT   DOMAIN      447    506       SUSHI 5.
FT   DOMAIN      509    568       SUSHI 6.
FT   DOMAIN      579    638       SUSHI 7.
FT   DOMAIN      641    700       SUSHI 8.
FT   DISULFID     60    158       BY SIMILARITY.
FT   DISULFID    131    150       BY SIMILARITY.
FT   DISULFID    163    174       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    185    194       BY SIMILARITY.
FT   DISULFID    200    244       BY SIMILARITY.
FT   DISULFID    230    257       BY SIMILARITY.
FT   DISULFID    262    306       BY SIMILARITY.
FT   DISULFID    292    319       BY SIMILARITY.
FT   DISULFID    324    368       BY SIMILARITY.
FT   DISULFID    354    381       BY SIMILARITY.
FT   DISULFID    386    430       BY SIMILARITY.
FT   DISULFID    416    443       BY SIMILARITY.
FT   DISULFID    448    492       BY SIMILARITY.
FT   DISULFID    478    505       BY SIMILARITY.
FT   DISULFID    510    554       BY SIMILARITY.
FT   DISULFID    540    567       BY SIMILARITY.
FT   DISULFID    580    624       BY SIMILARITY.
FT   DISULFID    610    637       BY SIMILARITY.
FT   DISULFID    642    686       BY SIMILARITY.
FT   DISULFID    672    699       BY SIMILARITY.
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    654    654       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    661    661       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       745    745       PALMITATE (BY SIMILARITY).
FT   SITE        756    759       ENDOCYTOSIS SIGNAL (PROBABLE).
FT   CONFLICT    721    724       A -> E, N (REF. 2).
SQ   SEQUENCE   768 AA;  83098 MW;  E5730C4D2F66E68 CRC64;

  Query Match             19.1%;  Score 143.5;  DB 1;  Length 768;
  Best Local Similarity   29.5%;  Pred. No. 7.5e-07;
  Matches   33;  Conservative   15;  Mismatches   49;  Indels   15;  Gaps    5;

Qy      29 RYTCSPSYRLIGEKAIFCISENQVHATWLKARVICESVNKTTSCSDPTVPY GGFMNKGS 86
           :: | | ||  |   : |    |   | :   ||:   |:| ' :| |     |
Db     289 KFECQPGYRAPGSNTLHCTGSGQ----WSEPLPTCEA---IACEPPEIPIHGSMDCVPS 340

Qy      87 KAPFPHGDSVTFTCKANFTMKGSKTVWCQANEMWGPTALPVCES---DFPL 134
             |  : | || |   | :||:  : |  :  |     | |||:   :||:
Db     341 TGTFGYNSSCTFLCAEGFVLKGNDAIQCADSGQWTAPA PECEALQCPEPPV 391


RESULT 20
CCPH_HSVSA
ID   CCPH_HSVSA      STANDARD;      PRT:   360 AA.
AC   Q01016;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Complement control protein homolog precursor (CCPH).
GN   4 OR CCPH
OS   Herpesvirus saimiri (strain 11).
```

```
OS   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10384;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333688; PubMed=1321287;
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
RN   [2]
RP   SIMILARITY TO CCP.
RX   MEDLINE=92260674; PubMed=1316492;
RA   Albrecht J.-C., Fleckenstein B.;
RT   "New member of the multigene family of complement control proteins in
RT   herpesvirus saimiri.";
RL   J. Virol. 66:4937-4940(1992).
CC   -!- ALTERNATIVE PRODUCTS: A MEMBRANE BOUND FORM AND A SECRETED FORM
CC       ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
CC       COMPLEMENT ACTIVATION (RCA).
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64446; CAA45626.1; -.
DR   EMBL; X64446; CAA45627.1; -.
DR   EMBL; X60283; CAA42823.1; -.
DR   EMBL; X60284; CAA42824.1; -.
DR   PIR; B42534; WMBE2E.
DR   PIR; A42534; WMBE1E.
DR   PIR; S24567; S24567.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4
KW   Signal; Repeat; Sushi; Transmembrane; Alternative splicing;
KW   Glycoprotein.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    360       COMPLEMENT CONTROL PROTEIN HOMOLOG.
FT   DOMAIN      83    143       SUSHI 1.
FT   DOMAIN     146    206       SUSHI 2.
FT   DOMAIN     209    265       SUSHI 3.
FT   TRANSMEM   328    350       POTENTIAL.
FT   DISULFID    84    125       BY SIMILARITY.
FT   DISULFID   111    142       BY SIMILARITY.
FT   DISULFID   147    191       BY SIMILARITY.
FT   DISULFID   175    205       BY SIMILARITY.
FT   DISULFID   210    252       BY SIMILARITY.
FT   DISULFID   238    264       BY SIMILARITY.
FT   CARBOHYD    36     36       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    39     39       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    46     46       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    72     72       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   155    155       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   294    294       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   289    302       RTCNGNCTISMPIQ - AECACPGSNYPISS (IN
FT                               SHORT ISOFORM).
FT   VARSPLIC   303    360       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   360 AA;  40006 MW;  627BA6C2ECD49660 CRC64;

  Query Match          16.9%;  Score 142.5;  DB 1;  Length 360;
  Best Local Similarity   27.2%;  Pred. No. 4.1e-07;
  Matches   37;  Conservative   18;  Mismatches   64;  Indels   17;  Gaps    5;

Qy       4 CDPPPEVKNAR     KIYYSLPIVPGTVLRYTCSPSYRLIGEKAIPCISENQVHATWDK 58
           |  :: : | |       ||      ||: ||| : | ||:     .: :      |
Db      84 CSTPQDLLNGRYTVTGNLYY----GSVITYTCNSGYSLIGSSIISCLLKGQGRVDWTP 137

Qy      59 APPICESVNETSCSDDIVIAGFMNKCSKAFKQGDSVIFVC  KANFIMKGSKIVWCQA 116
           '||: : |   | |  |       :   | : : |:|:::|  :  :: | :  :
Db     138 RPPICD-IKE---CKPPPQIANGTHTNVRDFYYYLDTVTYSCNEGFELLLPGSSELCSE 193

Qy     117 NEMWGPTALPVCESDF 132
             | |     ||  |
Db     194 TGSWVPNGEIKCEFIF 209


RESULT 21
DAF_HUMAN
ID   DAF_HUMAN      STANDARD;      PRT;   381 AA.
AC   P08174; P09679; P78361;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Complement decay-accelerating factor precursor (CD55 antigen).
GN   DAF OR CR1 OR CD55.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=87115845; PubMed=2433596;
RA   Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
RA   Nussenzweig V.;
RT   "Cloning of decay-accelerating factor suggests novel use of splicing
RT   to generate two proteins.";
RL   Nature 325:545-549(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervix;
RA   Stransberg R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 6-381 FROM N.A. (ISOFORM 2).
RX   MEDLINE=87175602; PubMed=2436222;
RA   Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
RA   Leykam J.F., Atkinson J.P., Tykocinski M.L.;
RT   "Cloning and characterization of cDNAs encoding the complete sequence
RT   of decay accelerating factor of human complement.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
RN   [4]
RP   SEQUENCE OF 35-381 FROM N.A. (ISOFORM 2).
RC   TISSUE=Hippocampus;
RA   Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
RT   "Decay-acceleration factor (DAF; CD-55) in the brain of Alzheimer's
RT   disease patients.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-100 FROM N.A.
RX   MEDLINE=91271256; PubMed=1711208;
RA   Ewulonu U.K., Ravi L., Medof M.E.;
RT   "Characterization of the decay-accelerating factor gene promoter
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
RN   [6]
RP   SEQUENCE OF 35-46.
RC   TISSUE=Urine;
RX   MEDLINE=91291869; PubMed=1712233;
RA   Nakano Y., Sugita Y., Ishikawa Y., Choi N. H., Tobe T., Tomita M.;
RT   "Isolation of two forms of decay-accelerating factor (DAF) from human
RT   urine.";
RL   Biochim. Biophys. Acta 1074:326-330(1991).
RN   [7]
RP   GPI-ANCHOR.
RX   MEDLINE=91093248; PubMed=1824699;
RA   Moran P., Raab H., Kohr W.J., Caras I.W.;
RT   "Glycophospholipid membrane anchor attachment. Molecular analysis of
```

```
RL   the cleavage/attachment site.";
RL   J. Biol. Chem. 266:1250-1257(1991).
RN   [8]
RP   DISULFIDE BONDS IN SUSHI DOMAINS
RX   MEDLINE=92305084; PubMed=1377029;
RA   Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
RT   "Complete determination of disulfide bonds localized within the short
RT   consensus repeat units of decay accelerating factor (CD55 antigen).";
RL   Biochim. Biophys. Acta 1116:235-240(1992).
RN   [9]
RP   FUNCTION AS A ECHOVIRUS RECEPTOR.
RX   MEDLINE=95045899; PubMed=7525274;
RA   Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
RA   Almond J.W.;
RT   "Decay accelerating factor CD55 is identified as the receptor for
RT   echovirus 7 using CELICS, a rapid immuno-focal cloning method.";
RL   EMBO J. 13:5070-5074(1994).
RN   [10]
RP   VARIANT BLOOD GROUP DR(A-).
RX   MEDLINE=94325524; PubMed=7519480;
RA   Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
RA   Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
RT   "Molecular basis of reduced or absent expression of
RT   decay-accelerating factor in Cromer blood group phenotypes.";
RL   Blood 84:1276-1282(1994).
CC   -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC       CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC       C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC       INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
CC       INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC       FACTOR B TO ENZYMATICALLY ACTIVE C2A AND BB AND THEREBY PREVENTS
CC       THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
CC       THE COMPLEMENT CASCADE.
CC   -!- FUNCTION: ALSO ACTS AS THE RECEPTOR FOR ECHOVIRUS 7 AND RELATED
CC       VIRUSES (ECHOVIRUSES 14, 21, 29 AND 33).
CC   -!- SUBUNIT: MONOMER (MAJOR FORM) AND NON-DISULFIDE-LINKED, COVALENT
CC       HOMODIMER (MINOR FORM).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/DAF-1 AND 2/DAF-2 (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON THE PLASMA MEMBRANES OF ALL CELL
CC       TYPES THAT ARE IN INTIMATE CONTACT WITH PLASMA COMPLEMENT
CC       PROTEINS. IT IS ALSO FOUND ON THE SURFACES OF EPITHELIAL CELLS
CC       LINING EXTRACELLULAR COMPARTMENTS, AND VARIANTS OF THE MOLECULE
CC       ARE PRESENT IN BODY FLUIDS AND IN EXTRACELLULAR MATRIX.
CC   -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC   -!- PTM: THE SER/THR-RICH DOMAIN IS HEAVILY O-GLYCOSYLATED.
CC   -!- POLYMORPHISM: DAF IS RESPONSIBLE FOR THE CROMER BLOOD GROUP
CC       SYSTEM. IT CONSISTS OF AT LEAST SEVEN HIGH-INCIDENCE (CR(A),
CC       TC(A), DR(A), ES(A), WES(B), UMC, AND IFC) AND LOW-INCIDENCE
CC       (TC(B), TC(C), AND WES(A)) ANTIGENS THAT RESIDE ON DAF. IN THE
CC       CROMER PHENOTYPES DR(A-) AND INAB THERE IS REDUCED OR ABSENT
CC       EXPRESSION OF DAF, RESPECTIVELY. IN THE CASE OF THE DR(A-)
CC       PHENOTYPE, A SINGLE NUCLEOTIDE SUBSTITUTION WITHIN EXON 5 ACCOUNTS
CC       FOR TWO CHANGES: A SIMPLE AMINO ACID SUBSTITUTION THAT IS THE
CC       BASIS OF THE ANTIGENIC VARIATION, AND AN ALTERNATIVE SPLICING
CC       EVENT THAT UNDERLIES THE DECREASED EXPRESSION OF DAF IN THIS
CC       PHENOTYPE.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD quide CD55 entry;
CC       WWW-"http://www.ncbi.nlm.nih.gov/prow/cd/cd55.htm".
CC   ---------- ----------- ----------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------- ----------- ----------------------------------------------------
DR   EMBL; M31516; AAA52169.1; -.
DR   EMBL; M30142; AAA52168.1; -.
DR   EMBL; BC001288; AAH01288.1; -.
DR   EMBL; M15799; AAA52167.1; -.
DR   EMBL; U88576; AAB48622.1; -.
DR   EMBL; M64653; AAA52170.1; -.
DR   EMBL; M64356; AAA52170.1; JOINED.
DR   EMBL; S72858; AAC60633.1; -.
DR   PIR; B26359; B26359.
DR   PIR; A26359; A26359.
DR   PIR; S16187; S16187.
DR   PIR; A39101; A39101.
DR   PIR; S23138; S23138.
DR   HSSP; P08603; 1HCC.
DR   MIM; 125240; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW   Alternative splicing; GPI anchor; Signal; Sushi; Polymorphism;
KW   Blood group antigen.
FT   SIGNAL        1     34
FT   CHAIN        35    353       COMPLEMENT DECAY-ACCELERATING FACTOR.
FT   PROPEP      354    381       REMOVED IN MATURE FORM.
FT   DOMAIN       35     95       SUSHI 1.
FT   DOMAIN       97    159       SUSHI 2.
FT   DOMAIN      162    221       SUSHI 3.
FT   DOMAIN      224    284       SUSHI 4.
FT   DOMAIN      287    356       SER/THR-RICH.
FT   DISULFID     36     81
FT   DISULFID     65     94
FT   DISULFID     98    145
FT   DISULFID    129    158
FT   DISULFID    163    204
FT   DISULFID    190    220
FT   DISULFID    225    267
FT   DISULFID    253    283
FT   CARBOHYD     95     95       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       353    353       GPI-ANCHOR.
FT   VARSPLIC    362    381       HTCFTLTGLGTLVTMGLLT -> SRPVTQAGMRWCTRSSL
FT                                QSRTPGPKESPHPSTPSSWYYRAHVPHVDREAWDASNHCLA
FT                                DLAKEELPEKYTQVYRLFLVS (IN ISOFORM 1).
FT   VARIANT      52     52       R -> L (IN TC(B) ANTIGEN).
FT                                /FTId=VAR_001997.
FT   VARIANT      52     52       R -> P (IN TC(C) ANTIGEN).
FT                                /FTId=VAR_001998.
FT   VARIANT      82     82       L -> R (IN WES(A) ANTIGEN).
FT                                /FTId=VAR_001999.
FT   VARIANT     199    199       S -> ... (IN DR(A-) ANTIGEN).
FT                                /FTId=VAR_002000.
FT   VARIANT     227    227       A -> P (IN CR(A-) ANTIGEN).
FT                                /FTId=VAR_002001.
FT   CONFLICT     80     80       T -> I (IN REF. 1, 2 AND 4).
FT   CONFLICT     85     85       S -> M (IN REF. 3).
FT   CONFLICT    187    187       S -> T (IN REF. 4).
FT   CONFLICT    297    297       Q -> H (IN REF. 4).
SQ   SEQUENCE   381 AA;  41388 MW;  2913BEEB6B4B865E CRC64;

  Query Match             18.9%;  Score 143;  DB 1;  Length 381;
  Best Local Similarity   29.9%;  Pred. No. 4.9e-07;
  Matches   40;  Conservative   24;  Mismatches   50;  Indels   20;  Gaps    9;

Qy       3 SCDPPPEVKNARKPYYSLP--IVPGTVLKYTCKPSYRLIGEKAIFCITSENQVHATWDKA 59
           ||    :::| :     .|  |.| . .. .  |:| |  : || || : |   |
Db     162 SCPNIPGEIRNGQ---IDVPGILFGATISFSCNTGYK FGSTSSFCLISGSSVQ--WSDP 216

Qy      60 PPICESVNKTISC-SDPIVPGGFMNKGSKAPLKHGDSVTFTCKANFTMKGSKTVWCQAN 117
            | |    : | | : | :  | : :| :   :  |||: |   ||| |  :::| |
Db     217 LPEC ---REIYCPAPPQIDNGII-QGERDHYGYRQSVTYACNKGFTMIGEHSIYCTVNN 271

Qy     118 ---EMWGPTALPVC 128
               |  ||   | |
```

```
Db    272 DEGIWSGIPS  PRT  283


RESULT 22
LEM2_BOVIN
ID   LEM2_BOVIN      STANDARD;      PRT;   485 AA.
AC   P98107;
DT   01-FEB-1996 (Rel. 34, Created)
DT   01-FEB-1996 (Rel. 34, Last sequence update)
DT   01-FEB-1996 (Rel. 34, Last annotation update)
DE   E selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=98382547; PubMed=9660465;
RA   Nguyen M., Strubel N.A., Bischoff J.;
RT   "A role for sialyl Lewis X/A glycoconjugates in capillary
RT   morphogenesis.";
RL   Nature 365:267-269(1993).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- FUNCTION: INVOLVED IN CAPILLARY MORPHOGENESIS BY BINDING TO A
CC       BOVINE CAPILLARY ENDOTHELIAL (BCE) CELL SIALYL LEWIS X AND/OR
CC       SIALYL LEWIS A CONTAINING LIGAND (PROBABLY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS; BOVINE E LECTIN LACKS
CC       THE HUMAN SUSHI 4 AND 5 EQUIVALENTS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; L12049; AAA02991.1; -.
DR   HSSI; PI0581; IESL.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 4.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 4.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL      1     22       BY SIMILARITY.
FT   CHAIN      23    485       E SELECTIN.
FT   DOMAIN     23    430       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   431   453       POTENTIAL.
FT   DOMAIN     454   485       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      39    139       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN     140   176       EGF-LIKE.
FT   DOMAIN     180   238       SUSHI 1.
FT   DOMAIN     241   300       SUSHI 2.
FT   DOMAIN     303   363       SUSHI 3.
FT   DOMAIN     366   422       SUSHI 4.
FT   DISULFID    41   139       BY SIMILARITY.
FT   DISULFID   112   131       BY SIMILARITY.
FT   DISULFID   144   155       BY SIMILARITY.
FT   DISULFID   149   164       BY SIMILARITY.
FT   DISULFID   166   175       BY SIMILARITY.
FT   DISULFID   181   224       BY SIMILARITY.
FT   DISULFID   210   237       BY SIMILARITY.
FT   DISULFID   242   286       BY SIMILARITY.
FT   DISULFID   272   299       BY SIMILARITY.
FT   DISULFID   304   349       BY SIMILARITY.
FT   DISULFID   335   362       BY SIMILARITY.
FT   DISULFID   367   408       BY SIMILARITY.
FT   DISULFID   394   421       BY SIMILARITY.
FT   CARBOHYD    61    61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    79    79       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    88    88       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   161   161       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   203   203       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   265   265       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   312   312       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   416   416       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   479   479       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   401   401       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   485 AA;  53200 MW;  AE931C9B521E3904 CRC64;

  Query Match          18.9%;  Score 142;  DB 1;  Length 485;
  Best Local Similarity  29.8%;  Pred. No. 6.4e-07;
  Matches   42;  Conservative   14;  Mismatches   55;  Indels   30;  Gaps    7;

Qy       2 ISCDPPPE----VKNARKPYYSLPIVPGIVLRYPCSPSYRELGEKAIFCTSENQVHAIWD 57
           :||       | |     : |||    |   : |   |:| | : : | |        ||
Db     240 VKCDALSNLDNGVVNCSPNEGSLPW  NTTCTIEPQQEGYKLTGPQHLQCISS    GIWD 293

Qy      58 KAPFCTESV---- ----NKTISCSHPIVPCGEMNKGSKAPFRHCDSVTFDCKANFIMKG 108
             | |::|          | |::||  :| | | |  |           | |:|   |||:|
Db     294 NKQFCKAVSCAAISHPQNGIVNCSHSVV GDPAFK          SSCHFDAEGFTLQG 343

Qy     109 SKTVWTQANEMWGPTALPVCE 129
           | | |   |    :||||
Db     344 -ESVVT AGTQW TQSVEVT 363


RESULT 23
LEM2_RAT
ID   LEM2_RAT        STANDARD;      PRT;   549 AA.
AC   P98105;
DT   01-FEB-1996 (Rel. 34, Created)
DT   01-FEB-1996 (Rel. 34, Last sequence update)
DT   01-FEB-1996 (Rel. 34, Last annotation update)
DE   E selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE OR ELAM 1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Rosenbloom C.L., Anchampach J.A., Anderson D.C., Manning A.M.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
```

```
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC     !  SUBCELLULAR LOCATION: Type I membrane protein.
CC     !  SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC     !  SIMILARITY: CONTAINS 1 C TYPE LECTIN FAMILY DOMAIN.
CC     !  SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC     !  SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L25527; AAA41113.1; -.
DR   HSSP; P16581; 1KJA.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    549       E-SELECTIN.
FT   DOMAIN       22    494       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    495    516       POTENTIAL.
FT   DOMAIN      517    549       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       38    138       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      139    175       EGF-LIKE.
FT   DOMAIN      179    239       SUSHI 1.
FT   DOMAIN      242    301       SUSHI 2.
FT   DOMAIN      304    364       SUSHI 3.
FT   DOMAIN      367    427       SUSHI 4.
FT   DOMAIN      430    486       SUSHI 5.
FT   DISULFID     40    148       BY SIMILARITY.
FT   DISULFID    111    130       BY SIMILARITY.
FT   DISULFID    143    154       BY SIMILARITY.
FT   DISULFID    148    163       BY SIMILARITY.
FT   DISULFID    165    174       BY SIMILARITY.
FT   DISULFID    180    225       BY SIMILARITY.
FT   DISULFID    210    238       BY SIMILARITY.
FT   DISULFID    243    287       BY SIMILARITY.
FT   DISULFID    273    300       BY SIMILARITY.
FT   DISULFID    305    350       BY SIMILARITY.
FT   DISULFID    336    363       BY SIMILARITY.
FT   DISULFID    368    413       BY SIMILARITY.
FT   DISULFID    399    426       BY SIMILARITY.
FT   DISULFID    431    472       BY SIMILARITY.
FT   DISULFID    458    485       BY SIMILARITY.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    145    145       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    192    192       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    313    313       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    320    320       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    441    441       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    465    465       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   549 AA;  60079 MW;  85CEECDB7B0144C8 CRC64;
```

```
  Query Match             18.6%;  Score 140;  DB 1;  Length 549;
  Best Local Similarity   29.6%;  Pred. No. 1.2e-06;
  Matches   34;  Conservative   13;  Mismatches   58;  Indels   10;  Gaps    4;

Qy      16 PYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPPICESVNKTISCSDP 75
           |  |     |: :::|   : | | : |   :    ||   | |  |     ||
Db     383 PSASAPFQSGSSCKFSCDEGFELKGSRRLQCFRGE-- WDSEKPICAGV--- QCSSI 434

Qy      76 IVPGGYMNKGSKAPFPHGDSVTFTCKANFTMKGSKTVWCQANEMWGPTALPVCES 130
             :||  ||     |   |    |||   :|: || : | |   |    || |||
Db     435 DLPGKFMNMSCSGPAVFGIVCEFTCPEGWILNGSSILICGATGRWS AMLPTCEA 487


RESULT 24
LEM2_HUMAN
ID   LEM2_HUMAN     STANDARD;      PRT:   610 AA.
AC   P16581; P15111;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE OR ELAM1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90175359; PubMed=1689848;
RA   Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,
RA   Pasek M., Pittack C., Tizard R., Goeli S., McCarthy K., Hopple S.,
RA   Lobb R.;
RT   "Endothelial leukocyte adhesion molecule 1: direct expression cloning
RT   and functional interactions.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89162047; PubMed=2466335;
RA   Bevilacqua M.P., Stengelin S., Gimbrone M.A. Jr., Seed B.;
RT   "Endothelial leukocyte adhesion molecule 1: an inducible receptor for
RT   neutrophils related to complement regulatory proteins and lectins.";
RL   Science 243:1160-1165(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91115870; PubMed=1703529;
RA   Collins T., Williams A., Johnston G.A., Kim J., Eddy R., Shows T.,
RA   Gimbrone M.A. Jr., Bevilacqua M.P.;
RT   "Structure and chromosomal location of the gene for endothelial-
RT   leukocyte adhesion molecule 1.";
RL   J. Biol. Chem. 266:2466-2473(1991).
RN   [4]
RP   LIGAND.
RX   MEDLINE=91068005; PubMed=1701274;
RA   Phillips M.L., Nudelman E., Gaeta F.C., Perez M., Singhal A.K.,
RA   Hakomori S., Paulson J.C.;
RT   "ELAM-1 mediates cell adhesion by recognition of a carbohydrate
RT   ligand, sialyl-Lex.";
RL   Science 250:1130-1132(1990).
RN   [5]
RP   3D-STRUCTURE MODELING OF LECTIN DOMAIN.
RX   MEDLINE=94202275; PubMed=7681016;
RA   Mills A.;
RT   "Modelling the carbohydrate recognition domain of human E-selectin.";
RL   FEBS Lett. 319:5-11(1993).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.
RX   MEDLINE=94150646; PubMed=7509040;
RA   Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,
RA   Huang K.-S., Presky D.H., Familletti P.C., Wolitzky B.A., Burns D.K.;
RT   "Insight into E-selectin/ligand interaction from the crystal
RT   structure and mutagenesis of the lec/EGF domains.";
```

```
RL   Nature 367:532-538(1994).
RN   [7]
RP   VARIANT ARG 149.
RX   MEDLINE=95179107; PubMed=7534825;
RA   Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schattke S.,
RA   Schulte K.L., Glaser C., Kohde K., Baumann G., Speer A.;
RT   "E-selectin polymorphism and atherosclerosis: an association study.";
RL   Hum. Mol. Genet. 3:1935-1937(1994).
RN   [8]
RP   VARIANT ARG 149.
RX   MEDLINE=99134508; PubMed=9933738;
RA   Ye S.Q., Usher D., Virtti D., Zhang L.Q., Yochim S.E., Gupta R.;
RT   "A PstI polymorphism detects the mutation of serine 128 to arginine in
RT   the E62E gene--a risk factor for coronary artery disease.";
RL   J. Biomed. Sci. 6:18-21(1999).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A
CC       HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY
CC       HIGHER MUTATION FREQUENCY (ARG 149) IS OBSERVED IN PATIENTS WITH
CC       ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN
CC       UNSELECTED POPULATION (SER 149).
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD62E entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/p_62e.htm"
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M30940; AAA52477.1; -.
DR   EMBL; M61893; AAA52475.1; -.
DR   EMBL; M61895; AAA52475.1; JOINED.
DR   EMBL; M61887; AAA52475.1; JOINED.
DR   EMBL; M61888; AAA52475.1; JOINED.
DR   EMBL; M61890; AAA52475.1; JOINED.
DR   EMBL; M61891; AAA52475.1; JOINED.
DR   EMBL; M61892; AAA52475.1; JOINED.
DR   EMBL; M24736; AAA52476.1; -.
DR   PIR; A32606; A32606.
DR   PIR; A35046; A35046.
DR   PIR; A38615; A38615.
DR   PDB; 1ESL; 31-AUG-94.
DR   PDB; 1KJA; 03-APR-96.
DR   MIM; 131210; -.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002496; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00434; SELECTIN.
DR   SMART; SM00032; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat; Polymorphism; 3D-structure.

FT   SIGNAL        1     21
FT   CHAIN        22    610       E-SELECTIN.
FT   DOMAIN       22    556       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     557   578       POTENTIAL.
FT   DOMAIN       579   610       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        38    138       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN       139   175       EGF-LIKE.
FT   DOMAIN       179   238       SUSHI 1.
FT   DOMAIN       241   300       SUSHI 2.
FT   DOMAIN       303   363       SUSHI 3.
FT   DOMAIN       366   426       SUSHI 4.
FT   DOMAIN       429   489       SUSHI 5.
FT   DOMAIN       492   548       SUSHI 6.
FT   DISULFID      40    138
FT   DISULFID     111   130
FT   DISULFID     143   154
FT   DISULFID     148   163
FT   DISULFID     165   174
FT   DISULFID     180   224       BY SIMILARITY.
FT   DISULFID     210   237       BY SIMILARITY.
FT   DISULFID     242   286       BY SIMILARITY.
FT   DISULFID     272   299       BY SIMILARITY.
FT   DISULFID     304   349       BY SIMILARITY.
FT   DISULFID     335   362       BY SIMILARITY.
FT   DISULFID     367   412       BY SIMILARITY.
FT   DISULFID     398   425       BY SIMILARITY.
FT   DISULFID     430   475       BY SIMILARITY.
FT   DISULFID     461   488       BY SIMILARITY.
FT   DISULFID     493   544       BY SIMILARITY.
FT   DISULFID     520   547       BY SIMILARITY.
FT   CARBOHYD      25    25        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      33    33        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     160   160       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     179   179       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     199   199       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     203   203       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     265   265       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     312   312       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     332   332       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     503   503       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     527   527       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT      130   130       C -> W (IN DBSNP:5360).
FT                                /FTId=VAR_011790.
FT   VARIANT      149   149       S -> R (ASSOCIATED WITH A RISK FACTOR FOR
FT                                CAD).
FT                                /FTId=VAR_004191.
FT   VARIANT      295   295       E -> K (IN DBSNP:5364).
FT                                /FTId=VAR_011791.
FT   VARIANT      421   421       E -> Q (IN DBSNP:5366).
FT                                /FTId=VAR_011792.
FT   VARIANT      468   468       H -> Y (IN DBSNP:5368).
FT                                /FTId=VAR_011793.
FT   VARIANT      575   575       L -> F (IN DBSNP:5369).
FT                                /FTId=VAR_011794.
SQ   SEQUENCE   610 AA;  66655 MW;  7D43E3C0D122F229 CRC64;

  Query Match             18.5%;  Score 139;  DB 1;  Length 610;
  Best Local Similarity   31.5%;  Pred. No. 1.7e-06;
  Matches   44;  Conservative   18;  Mismatches   86;  Indels   20;  Gaps    6;

Qy      30 YTCSPSYRLLGEKAIFCTSENQVHAIWDKAPLICKSVNKTISCPLPVGFMN  KQSK 87
           :||   : | |   : | :: |    | :  |:||:   | : |:|    :|||    :
Db     333 FTCEEGFMLQGPAQVECTTQGQ---WTQQIPVCEAFQCT ALSNP  ERGYMNCLPSAS 385

Qy      88 APFRHGDSVIFTCKANFTMKGSKTVWCQANEMWGPTA     LIVCES 130
             |:: | |:|:  | :|||| : |     |||       | |||
Db     386 GSFRYGSSCEFSCEQGFVLKGSKRLQC   -  GPTGEWDNEKPTCEA 427


RESULT 25
F13B_HUMAN
ID   F13B_HUMAN     STANDARD;      PRT;   661 AA.
AC   P05160;
```

```
DT   14-AUG-1987 (Rel. 05, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein-
DE   glutamine gamma-glutamyltransferase B chain) (Transglutaminase B
DE   chain).
GN   F13B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91105054; PubMed=2271707;
RA   Bottenus R.E., Ichinose A., Davie E.W.;
RT   "Nucleotide sequence of the gene for the b subunit of human factor
RT   XIII.";
RL   Biochemistry 29:11195-11209(1990).
RN   [2]
RP   SEQUENCE OF 2-661 FROM N.A.
RX   MEDLINE=87026535; PubMed=4021194;
RA   Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;
RT   "Amino acid sequence of the b subunit of human factor XIII, a protein
RT   composed of ten repetitive segments.";
RL   Biochemistry 25:4633-4638(1986).
RN   [3]
RP   REVISIONS.
RA   Ichinose A.;
RL   Submitted (FEB-1987) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-20 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90251467; PubMed=2339867;
RA   Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;
RT   "Complete cDNA sequence encoding the B subunit of human factor XIII.";
RL   Nucleic Acids Res. 18:2817-2817(1990).
RN   [5]
RP   VARIANT PHE 450.
RX   MEDLINE=93313189; PubMed=8324218;
RA   Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;
RT   "Two genetic defects in a patient with complete deficiency of the b-
RT   subunit for coagulation factor XIII.";
RL   Blood 82:145-150(1993).
CC   -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC   -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC   -!- DISEASE: A DEFICIENCY IN FA13 CAN RESULT IN A LIFELONG BLEEDING
CC       TENDENCY, DEFECTIVE WOUND HEALING, AND HABITUAL ABORTION.
CC   -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64554; AAA51821.1; ALT_SEQ.
DR   EMBL; M14057; AAA88042.1; -.
DR   EMBL; X51824; CAA36123.1; -.
DR   PIR; A23830; A23830.
DR   PIR; A46397; A46397.
DR   PIR; S09980; S09980.
DR   HSSP; P08603; 1HFI.
DR   MIM; 134580; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi; Disease mutation.
FT   SIGNAL       1     20
FT   CHAIN        21    661       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN       24     88       SUSHI 1.
FT   DOMAIN       90    147       SUSHI 2.
FT   DOMAIN      152    209       SUSHI 3.
FT   DOMAIN      212    268       SUSHI 4.
FT   DOMAIN      273    328       SUSHI 5.
FT   DOMAIN      335    390       SUSHI 6.
FT   DOMAIN      395    451       SUSHI 7.
FT   DOMAIN      453    516       SUSHI 8.
FT   DOMAIN      523    579       SUSHI 9.
FT   DOMAIN      581    647       SUSHI 10.
FT   DISULFID     25     76       BY SIMILARITY.
FT   DISULFID     59     87       BY SIMILARITY.
FT   DISULFID     91    135       BY SIMILARITY.
FT   DISULFID    118    146       BY SIMILARITY.
FT   DISULFID    153    197       BY SIMILARITY.
FT   DISULFID    180    208       BY SIMILARITY.
FT   DISULFID    213    255       BY SIMILARITY.
FT   DISULFID    241    267       BY SIMILARITY.
FT   DISULFID    274    316       BY SIMILARITY.
FT   DISULFID    302    327       BY SIMILARITY.
FT   DISULFID    336    378       BY SIMILARITY.
FT   DISULFID    364    389       BY SIMILARITY.
FT   DISULFID    396    439       BY SIMILARITY.
FT   DISULFID    425    450       BY SIMILARITY.
FT   DISULFID    454    505       BY SIMILARITY.
FT   DISULFID    486    515       BY SIMILARITY.
FT   DISULFID    524    567       BY SIMILARITY.
FT   DISULFID    553    578       BY SIMILARITY.
FT   DISULFID    582    636       BY SIMILARITY.
FT   DISULFID    616    646       BY SIMILARITY.
FT   CARBOHYD    162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        617    619       CELL ATTACHMENT SITE.
FT   VARIANT     450    450       C -> F (IN F13B DEFICIENCY).
FT                                /FTId=VAR_007475.
SQ   SEQUENCE   661 AA;  75491 MW;  57A2FB4656085/F2 CRC64;

  Query Match           18.2%;  Score 137;  DB 1;  Length 661;
  Best Local Similarity 27.9%;  Pred. No. 2.9e-06;
  Matches   36;  Conservative   19;  Mismatches   62;  Indels   12;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGPKAIECTSENQVHAIWDKAP 60
           :::|: || ::|     :|     |  :   |  | | |  | |           |   |
Db     333 KVACEEPPFIENGAANLHSKIYYNGDKVTYACKSGYLLEGSNEIIC-----NRGKW-ILP 386

Qy      61 PICESVNKTISCS-DPIVPGGFMNKGSKAPFREGDSVTFTCKANFTMKGSKSVWCQANEM 119
           | |   |:  :|   |:| | :  |  |  :  | || : |   : ::|||   |:  :
Db     387 PECVENNE--NCKHPPVVMNGAVADGILASYAIGSSVFYRCNEYYLLRGSKISRCEQGKW 444

Qy     120 WGPTALPVC 128
             |   | |
Db     445 SSP---PVC 450


RESULT 26
F13B_MOUSE
ID   F13B_MOUSE     STANDARD;      PRT;   668 AA.
AC   Q07968;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Coagulation factor XIII B chain precursor (EC 2.3.2.13) (Protein
DE   glutamine gamma-glutamyltransferase B chain) (Transglutaminase B
DE   chain).
GN   F13B OR CF13B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.D2/OSN; TISSUE=Liver;
```

```
RX   MEDLINE=93224141; PubMed=8460048;
RA   Nonaka M., Matsuda Y., Shiroishi T., Moriwaki K., Nonaka M.;
RA   Nat Genome Sakai S.;
RT   "Molecular cloning of the b subunit of mouse coagulation factor XIII
RT   and assignment of the gene to chromosome 1, close evolutionary
RT   relationship to complement factor H.";
RL   Genomics 15:545-542(1994).
CC    !  FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,
CC       BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE
CC       OF TRANSGLUTAMINASE FORMATION BY THROMBIN.
CC    !  SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.
CC    !  TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.
CC    !  SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.
CC   ---------- ------ ----- ---------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------ ----- ---------
DR   EMBL; D16071; BAA00063.1; -.
DR   PIR; A46013; A46013.
DR   HSSP; P08603; 1BE1.
DR   MGD; MGI:98379; F13B.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 10.
DR   SMART; SM00032; CCP; 10.
KW   Transferase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;
KW   Sushi.
FT   SIGNAL       1     20       BY SIMILARITY.
FT   CHAIN       21    668       COAGULATION FACTOR XIII B CHAIN.
FT   DOMAIN      24     88       SUSHI 1.
FT   DOMAIN      90    147       SUSHI 2.
FT   DOMAIN     152    209       SUSHI 3.
FT   DOMAIN     212    268       SUSHI 4.
FT   DOMAIN     273    328       SUSHI 5.
FT   DOMAIN     335    390       SUSHI 6.
FT   DOMAIN     395    451       SUSHI 7.
FT   DOMAIN     453    516       SUSHI 8.
FT   DOMAIN     523    579       SUSHI 9.
FT   DOMAIN     581    647       SUSHI 10.
FT   DISULFID    25     76       BY SIMILARITY.
FT   DISULFID    59     87       BY SIMILARITY.
FT   DISULFID    91    145       BY SIMILARITY.
FT   DISULFID   118    146       BY SIMILARITY.
FT   DISULFID   153    197       BY SIMILARITY.
FT   DISULFID   180    208       BY SIMILARITY.
FT   DISULFID   214    255       BY SIMILARITY.
FT   DISULFID   241    267       BY SIMILARITY.
FT   DISULFID   274    316       BY SIMILARITY.
FT   DISULFID   302    327       BY SIMILARITY.
FT   DISULFID   336    378       BY SIMILARITY.
FT   DISULFID   364    389       BY SIMILARITY.
FT   DISULFID   396    449       BY SIMILARITY.
FT   DISULFID   425    450       BY SIMILARITY.
FT   DISULFID   454    505       BY SIMILARITY.
FT   DISULFID   486    515       BY SIMILARITY.
FT   DISULFID   524    567       BY SIMILARITY.
FT   DISULFID   553    578       BY SIMILARITY.
FT   DISULFID   582    636       BY SIMILARITY.
FT   DISULFID   616    646       BY SIMILARITY.
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   545    545       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   668 AA;  76078 MW;  80BC9E00A9E53FA6 CRC64;

  Query Match          18.0%;  Score 135;  DB 1;  Length 668;
  Best Local Similarity   27.6%;  Pred. No. 4.6e-06;
  Matches   47;  Conservative   18;  Mismatches   57;  Indels   22;  Gaps    6;

Qy       1 EISCDPPPEVKNA      RKFYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAT 55
           ::::: || |:|       : |||     |  : | |   | | |   | |

Db     333 KVACEQPPSVENGVAHPHSEIYYS     GDKVIYRCGGGYSLRGSSTIT      NRGR 382

Qy      56 WDFAPPLPFSVNETISCS-DPEIPNGTFMNRCSFAPRPIPSVIIS-FANLIMEGCETVWC 114
           |  | |  |:    :|  | :  | :  | | :  | || : |   : ||||:  |
Db     383 WIIPPFCVFNIE---NCKPPPDIANGVVVDCLLASYIIGSSVRYSCNEYYILRGSETSRC 439

Qy     115 QANEMWGPTALPVC 128
           :    |   |||
Db     440 EQGAWSSP--PVC 450


RESULT 27
APOH_RAT
ID   APOH_RAT        STANDARD;       PRT;   297 AA.
AC   P26644;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2 glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI).
GN   APOH.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY;
RX   MEDLINE=89366680; PubMed=2771654;
RA   Aoyama Y., Chan Y.L., Wool I.G.;
RT   "The primary structure of rat beta-2 glycoprotein I.";
RL   Nucleic Acids Res. 17:6401-6401(1989).
CC    !- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------- ------ ----- ---------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------ ----- ---------
DR   EMBL; X15551; CAA33556.1; -.
DR   PIR; S05310; NBRT.
DR   HSSP; P02749; 1C1Z.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19       POTENTIAL.
FT   CHAIN       20    297       BETA-2 GLYCOPROTEIN I.
FT   DOMAIN      20     32       SUSHI 1 (INCOMPLETE).
FT   DOMAIN      35     90       SUSHI 2.
FT   DOMAIN      93    153       SUSHI 3.
FT   DOMAIN     156    213       SUSHI 4.
FT   DOMAIN     215    297       SUSHI-LIKE.
FT   DISULFID    36     76       BY SIMILARITY.
FT   DISULFID    62     89       BY SIMILARITY.
FT   DISULFID    94    140       BY SIMILARITY.
FT   DISULFID   126    152       BY SIMILARITY.
FT   DISULFID   157    200       BY SIMILARITY.
FT   DISULFID   186    212       BY SIMILARITY.
FT   DISULFID   216    267       BY SIMILARITY.
FT   DISULFID   252    277       BY SIMILARITY.
FT   DISULFID   259    297       BY SIMILARITY.
FT   CARBOHYD    69     69       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   114    114       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   135    135       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   CARBOHYD     145     145       N LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     205     205       N LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   297 AA;  33197 MW;  911209DA6C119D59 CRC64;

  Query Match             17.8%;  Score 134;  DB 1;  Length 297;
  Best Local Similarity   26.6%;  Pred. No. 2.4e-06;
  Matches   37;  Conservative   19;  Mismatches   65;  Indels   18;  Gaps    6;

Qy       2 ISCDPPPEVKNARKPYYSLPI---- VPGIVLRYTCSPSYRLIGEKAIFCISENQVHATW 56
           |:| |||  | |    |  :          ||: : | | : : |   : | :    | |
Db      92 ITCPPPPIPKFAALKEYKTSVGNSSFYQDIVV-FKCLPHFAMFGNDTVTCTA----HGNW 146

Qy      57 DKAPPICESVNKTISCSDPIVP-GGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQ 115
             : |   |      : |  |  |  ||:|  :|    :|   | |   : : | : | |
Db     147 TQLPECRE-----VKCPFPSRPDNGFVNYPAKPVLSYKDKAVFGCHETYKLDGPEEVECT 201

Qy     116 ANEMWGPTALPVCESDFPL 134
               |  :||| |::   |
Db     202 KTGNW  SALPSCKASCKL 218


RESULT 28
C4BB_RAT
ID   C4BB_RAT       STANDARD;      PRT;   258 AA.
AC   Q64515;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   C4B binding protein beta chain precursor.
GN   C4BPB.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY; TISSUE=Liver;
RX   MEDLINE=97166082; PubMed=9013975;
RA   Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT   "Molecular cloning of rat C4b binding protein alpha- and beta-chains:
RT   structural and functional relationships among human, bovine, rabbit,
RT   mouse, and rat proteins.";
RL   J. Immunol. 158:1315-1323(1997).
CC   -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC       ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC       (C4BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC       ALSO ACCELERATES THE DECRADATION OF THE C4BC2A COMPLEX (C3
CC       CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. IT
CC       INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID
CC       P COMPONENT.
CC   -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: TO C4BP ALPHA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; 250052; CAA90392.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
KW   Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL        1      15       BY SIMILARITY.
FT   CHAIN        16     258       C4B BINDING PROTEIN BETA CHAIN.
FT   DOMAIN       19      74       SUSHI 1 (ATYPICAL; LACK A CYS).
FT   DOMAIN       77     132       SUSHI 2.
FT   DOMAIN      135     189       SUSHI 3.
FT   DISULFID     46      73       BY SIMILARITY.
FT   DISULFID     78     118       BY SIMILARITY.
FT   DISULFID    136     176       BY SIMILARITY.
FT   DISULFID    162     188       BY SIMILARITY.
FT   DISULFID    217     217       INTERCHAIN (WITH ALPHA CHAIN)
FT                                 (POTENTIAL).
FT   CARBOHYD     27      27       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     67      67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89      89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     95      95       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114     114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    218     218       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   258 AA;  28641 MW;  636F9990BAB5DC1C CRC64;

  Query Match             17.8%;  Score 133.5;  DB 1;  Length 258;
  Best Local Similarity   26.1%;  Pred. No. 2.3e-06;
  Matches   35;  Conservative   17;  Mismatches   57;  Indels   25;  Gaps    6;

Qy       3 SCDPPPEVKNARKPYYSLPIVPGIVLR------ YTCSPSYRLIGEKAIFCISENQVHATW 56
           || |  | |:         :  |           |  | |:|:|::       :    |
Db      19 SCSEPPPVNNS--------VFVGKETEEQTLG YLCIKGYHLVGKKSLVFDPSKE---W 66

Qy      57 DKAPP CESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQA 116
           :   | |   :    | ||::  | :|  |          : | |   : :|||    |
Db      67 NSTLPECLLGH----CPDPVLENGKIN--SSGIDYNISGKIMFECNDGYILKGSNWSQCLE 120

Qy     117 NEMWGPTALPVCES 130
           :  |    ||:| |
Db     121 DRTWAP-PLPICRS 133


RESULT 29
LEM2_MOUSE
ID   LEM2_MOUSE     STANDARD;      PRT;   662 AA.
AC   Q00690;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE OR ELAM-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92283265; PubMed=1375914;
RA   Becker-Andre M., van Huijsduijnen R.H. Losberger C., Whelan J.,
RA   De amarter J.F.;
RT   "Murine endothelial leukocyte adhesion molecule 1 is a close
RT   structural and functional homologue of the human protein.";
RL   Eur. J. Biochem. 206:401-411(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92340571; PubMed=1378846;
RA   Weller A., Isenmann S., Vestweber D.;
RT   "Cloning of the mouse endothelial selectins. Expression of both E-
RT   and P-selectin is inducible by tumor necrosis factor alpha.";
RL   J. Biol. Chem. 267:15176-15183(1992).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstat on -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M80778; AAA37547.1; -.
DR   EMBL; M87862; AAA37577.1; ALT_INIT.
DR   HSSP; P16581; 1KJA.
DR   MGD; MGI:98227; Selp.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; Lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL       1     21       BY SIMILARITY.
FT   CHAIN       22    612       E-SELECTIN.
FT   DOMAIN      22    557       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    558   579       POTENTIAL.
FT   DOMAIN      580   612       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       38    138      C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      139    175      EGF-LIKE.
FT   DOMAIN      179    239      SUSHI 1.
FT   DOMAIN      242    301      SUSHI 2.
FT   DOMAIN      304    364      SUSHI 3.
FT   DOMAIN      367    427      SUSHI 4.
FT   DOMAIN      430    490      SUSHI 5.
FT   DOMAIN      493    549      SUSHI 6.
FT   DISULFID     40    138      BY SIMILARITY.
FT   DISULFID    111    130      BY SIMILARITY.
FT   DISULFID    143    154      BY SIMILARITY.
FT   DISULFID    148    163      BY SIMILARITY
FT   DISULFID    165    174      BY SIMILARITY.
FT   DISULFID    180    225      BY SIMILARITY.
FT   DISULFID    210    238      BY SIMILARITY.
FT   DISULFID    244    287      BY SIMILARITY.
FT   DISULFID    273    306      BY SIMILARITY.
FT   DISULFID    305    456      BY SIMILARITY.
FT   DISULFID    336    463      BY SIMILARITY.
FT   DISULFID    368    413      BY SIMILARITY.
FT   DISULFID    399    426      BY SIMILARITY.
FT   DISULFID    431    476      BY SIMILARITY.
FT   DISULFID    462    489      BY SIMILARITY.
FT   DISULFID    494    545      BY SIMILARITY.
FT   DISULFID    521    548      BY SIMILARITY.
FT   CARBOHYD     25     25      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    145    145      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    192    192      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    203    203      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    266    266      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    333    333      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    391    391      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    504    504      N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    528    528      N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   612 AA;  66749 MW;  86F05714F0E2C4D CRC64;

  Query Match             17.5%;  Score 131.5;  DB 1;  Length 612;
  Best Local Similarity   28.1%;  Pred. No. 9.4e-06;
  Matches   84;  Conservative   17,  Mismatches   41,  Indels   29,  Gaps    6,

Qy       24 PG     -IVLRYTCSPSYRLLGEKAIFCISENQVHATWDKAPPICESV          NKT 69
            ||      |   : |   || :|  :  : | |        ||   |  |:::|      |
Db      260 PGSYPWNTTCTFDCVEGYPPVGAQNICGTSS    GIWDNETPSCKAVTCDAIPQPQNGF 315

Qy       70 IGCSATFVPGETMNFGSKAFFECVETLDGFANWMFFIVW QANEMWGPTALPVCE 129
            :|||           :   :  |:   |  |||: :||:|   | | |   | |  :|||:
Db      316 VSCSH         STAGELAFK-SSCNFTCEQSFTLQGPAQVECSAQGQWTP QTPVCK 364

Qy      130 S 130
            :
Db      365 A 365


RESULT 30
CO2_HUMAN
ID   CO2_HUMAN       STANDARD;       PRT;   752 AA.
AC   P06681; Q13904;
DT   01-JAN-1988 (Rel. 06, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase).
GN   C2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=87127920; PubMed=2949737;
RA   Bentley D.R.;
RT   "Primary structure of human complement component C2. Homology to two
RT   unrelated protein families.";
RL   Biochem. J. 239:339-345(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93315833; PubMed=8326124;
RA   Ishii Y., Zhu Z.B., Macon K.J., Volanakis J.E.;
RT   "Structure of the human C2 gene.";
RL   J. Immunol. 151:170-174(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Swartzell S., Smith I.M., Spies T., Hood L.;
RT   "Sequence determination of 800 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 587-752 FROM N.A.
RX   MEDLINE=84144868; PubMed=6199794;
RA   Bentley D.R., Porter R.R.;
RT   "Isolation of cDNA clones for human complement component C2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:1212-1215(1984).
RN   [5]
RP   SEQUENCE OF 694-752 FROM N.A.
RX   MEDLINE=87102880; PubMed=3643061;
RA   Wu L., Morley B.J., Campbell R.D.;
RT   "Cell-specific expression of the human complement protein factor B
RT   gene: evidence for the role of two distinct 5' flanking elements.";
RL   Cell 48:333-342(1987).
RN   [6]
RP   SEQUENCE OF 137-171; 454-466 AND 574-717.
RX   MEDLINE=85038851; PubMed=6149575;
RA   Gagnon J.;
RT   "Structure and activation of complement components C2 and factor B.";
RL   Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:401-409(1984).
RN   [7]
RP   SEQUENCE OF 244-269.
RX   MEDLINE=83308518; PubMed=6555044;
RA   Parkes C., Gagnon J., Kerr M.A.;
RT   "The reaction of iodine and thiol blocking reagents with human
RT   complement components C2 and factor B. Purification and N-terminal
RT   amino acid sequence of a peptide from C2a containing a free thiol
```

```
RT   group.";
RL   Biochem. J. 213:201-209(1983).
RN   [8]
RP   VARIANTS C2D PHE-209 AND ARG-464.
RX   MEDLINE=96215049; PubMed=8621452;
RA   Wetsel R.A., Kulics J., Lokki M.L., Kiepiela P., Akama H.,
RA   Johnson C.A., Densen P., Colten H.R.;
RT   "Type II human complement C2 deficiency. Allele specific amino acid
RT   substitutions (Ser189 --> Phe; Gly444 --> Arg) cause impaired C2
RT   secretion.";
RL   J. Biol. Chem. 271:5824-5831(1996).
RN   [9]
RP   VARIANT C2D TYR-131.
RX   MEDLINE=98334005; PubMed=9670930;
RA   Zhu Z.B., Atkinson T.P., Volanakis J.E.,
RT   "A novel type II complement C2 deficiency allele in an African-
RT   American family.";
RL   J. Immunol. 161:578-584(1998).
CC   -!- FUNCTION: COMPONENT C2 WHICH IS PART OF THE CLASSICAL PATHWAY OF
CC       THE COMPLEMENT SYSTEM IS CLEAVED BY ACTIVATED FACTOR C1 INTO TWO
CC       FRAGMENTS: C2B AND C2A. C2A, A SERINE PROTEASE, THEN COMBINES WITH
CC       COMPLEMENT FACTOR 4B TO GENERATE THE C3 OR C5 CONVERTASE.
CC   -!- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND
CC       C4B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH
CC       CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.
CC   -!- DISEASE: DEFECTS IN C2 ARE THE CAUSE OF C2 DEFICIENCY (C2D). THIS
CC       IS AN AUTOSOMAL RECESIVE DISEASE. DEFICIENT INDIVIDUALS HAVE AN
CC       INCREASED INCIDENCE OF SLE AND SLE-LIKE SYNDROMES,
CC       GLOMERULONEPHRITIS, VASCULITIS AND PYOGENIC INFECTIONS. TYPE I C2D
CC       IS CHARACTERIZED BY COMPLETE LOSS OF THE PROTEIN WHILE TYPE II C2D
CC       IS CHARACTERIZED BY A SELECTIVE BLOCK IN C2 SECRETION.
CC   -!- MISCELLANEOUS: C2 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
CC       PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT FACTOR B.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M15082; AAA59624.1; -.
DR   EMBL; L09708; AAB97607.1; -.
DR   EMBL; L09706; AAB97607.1; JOINED.
DR   EMBL; L09707; AAB97607.1; JOINED.
DR   EMBL; AF019413; AAB67975.1; -.
DR   EMBL; X04481; CAA28169.1; -.
DR   PIR; A25290; C2HU.
DR   PIR; A05289; A05289.
DR   HSSP; P00734; 1VR1.
DR   MEROPS; S01.194; -.
DR   MIM; 217000; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 3.
DR   Pfam; PF00089; trypsin; 2.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 4.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;
KW   Signal; Repeat; Sushi; Disease mutation; Polymorphism.
FT   SIGNAL       1     20
FT   CHAIN       21    752       COMPLEMENT C2.
FT   CHAIN       21    243       COMPLEMENT C2B FRAGMENT.
FT   CHAIN      244    752       COMPLEMENT C2A FRAGMENT.
FT   DOMAIN      23     85       SUSHI 1.
FT   DOMAIN      88    145       SUSHI 2.
FT   DOMAIN     150    205       SUSHI 3.
FT   DOMAIN     254    452       VWFA.
FT   DOMAIN     466    752       SERINE PROTEASE.
FT   ACT_SITE   507    507       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   561    561       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   679    679       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID    24     64       BY SIMILARITY.
FT   DISULFID    51     84       BY SIMILARITY.
FT   DISULFID    89    131       BY SIMILARITY.
FT   DISULFID   117    144       BY SIMILARITY.
FT   DISULFID   151    191       BY SIMILARITY.
FT   DISULFID   177    204       BY SIMILARITY.
FT   CARBOHYD    29     29       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   112    112       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   290    290       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   333    333       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   467    467       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   471    471       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   621    621       N-LINKED (GLCNAC...).
FT   CARBOHYD   651    651       N-LINKED (GLCNAC...).
FT   VARIANT    131    131       C -> Y (IN C2D, TYPE II).
FT                               /FTId=VAR_008544.
FT   VARIANT    209    209       S -> F (IN C2D, TYPE II).
FT                               /FTId=VAR_008545.
FT   VARIANT    464    464       G -> R (IN C2D, TYPE II).
FT                               /FTId=VAR_008546.
FT   VARIANT    533    533       P -> L (IN DBSNP:1042664).
FT                               /FTId=VAR_011772.
SQ   SEQUENCE   752 AA;  83267 MW;  5A96A1CE700CF444 CRC64;

  Query Match             17.4%;  Score 130.5;  DB 1;  Length 752;
  Best Local Similarity   27.0%;  Pred. No. 1.5e-05;
  Matches   38;  Conservative   19;  Mismatches   61;  Indels   23;  Gaps    8;

Qy        2 ISCDPPPEVKNARKPYY-----SLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATW 56
            : |  |   :|     |     | ||: |  :   |   : | |     |   : :  |
Db       87 VRCPAPVSPENG---IYTPRLGSYPV--GGNVSFECEDGFILRGSPVRQC----RPNGMW 137

Qy       57 DKAPPICESVNKTISCSDPIVPGGFMNKGSKAIFYRHGDSVTFTCKANFTMKGSKIVWCQA 116
            |    :|:  |   | :| :  | :  | :   |||  | : | :|   : ||     |
Db      138 DGETAVCD--NGAGHCPNPGISLGAVRTGFT--EGHGDEVRYRCSSNLVLTGSSERECQG 193

Qy      117 NEMWGPTALPVCES----DFP 133
            | :|  |  |:|       |||
Db      194 NGVWSCTE-PICRQPYSYDFP 213


RESULT 31
CFAB_MOUSE
ID   CFAB_MOUSE     STANDARD;      PRT;   761 AA.
AC   P04186;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor B precursor (EC 3.4.21.47) (C3/C5 convertase).
GN   BF OR H2-BF.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91035430; PubMed=2229060;
RA   Ishikawa N., Nonaka M., Wetsel R.A., Colten H.R.;
RT   "Murine complement C2 and factor B genomic and cDNA cloning reveals
```

```
RT   different mechanisms for multiple transcripts of C2 and B.";
RL   J. Biol. Chem. 265:19041-19046(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Bowen L., Qin S., Lasky S.R., Loretz C., Dors M., Mahairas G.,
RA   Hood L.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RA   Bowen L., Madan A., Qin S., Ball J., Dahl T., James R., Dickhoff R.,
RA   Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;
RT   "Sequence of the mouse MHC class III region.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 285-761 FROM N.A.
RX   MEDLINE=84061941; PubMed=6089022;
RA   Sackstein R., Colten H.R., Woods D.E.;
RT   "Phylogenetic conservation of a class III major histocompatibility
RT   complex antigen, factor B. Isolation and nucleotide sequencing of
RT   mouse factor B cDNA clones.";
RL   J. Biol. Chem. 258:14693-14697(1983).
CC   -!- FUNCTION: FACTOR B WHICH IS PART OF THE ALTERNATE PATHWAY OF THE
CC       COMPLEMENT SYSTEM IS CLEAVED BY FACTOR D INTO 2 FRAGMENTS: BA AND
CC       BB. BB, A SERINE PROTEASE, THEN COMBINES WITH COMPLEMENT FACTOR 3B
CC       TO GENERATE THE C3 OR C5 CONVERTASE.
CC   -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha chain to yield C3a and
CC       C3b. Cleaves C5 in the alpha chain to yield C5a and C5b. Both
CC       cleavages take place at the C terminal of an arginine residue.
CC   -!- MISCELLANEOUS: FACTOR B IS A MAJOR HISTOCOMPATIBILITY COMPLEX
CC       CLASS III PROTEIN.
CC   -!- SIMILARITY: WITH COMPLEMENT C2.
CC   -!- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M60646; AAA37379.1; -.
DR   EMBL; M60629; AAA37379.1; JOINED.
DR   EMBL; M60630; AAA37379.1; JOINED.
DR   EMBL; M60631; AAA37379.1; JOINED.
DR   EMBL; M60632; AAA37379.1; JOINED.
DR   EMBL; M60633; AAA37379.1; JOINED.
DR   EMBL; M60634; AAA37379.1; JOINED.
DR   EMBL; M60635; AAA37379.1; JOINED.
DR   EMBL; M60636; AAA37379.1; JOINED.
DR   EMBL; M60637; AAA37379.1; JOINED.
DR   EMBL; M60638; AAA37379.1; JOINED.
DR   EMBL; M60639; AAA37379.1; JOINED.
DR   EMBL; M60640; AAA37379.1; JOINED.
DR   EMBL; M60641; AAA37379.1; JOINED.
DR   EMBL; M60642; AAA37379.1; JOINED.
DR   EMBL; M60643; AAA37379.1; JOINED.
DR   EMBL; M60644; AAA37379.1; JOINED.
DR   EMBL; M60645; AAA37379.1; JOINED.
DR   EMBL; AF049850; AAC05284.1; -.
DR   EMBL; AF109906; AAC84160.1; -.
DR   EMBL; K01496; AAA39549.1; -.
DR   EMBL; K01497; AAA39550.1; -.
DR   EMBL; K01498; AAA39551.1; -.
DR   EMBL; M57890; AAA64293.1; -.
DR   PIR; A38875; BBMS.
DR   HSSP; P00734; 2HNT.
DR   MEROPS; S01.196; -.
DR   MGI; MGI:105975; H2-Bf.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00084; sushi; 3.
DR   Pfam; PF00089; trypsin; 2.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Complement alternate pathway; Plasma; Hydrolase; Serine protease;
KW   Glycoprotein; Repeat; Sushi; Signal.
FT   SIGNAL       1     22       BY SIMILARITY.
FT   CHAIN       23    761       COMPLEMENT FACTOR B.
FT   CHAIN       23    256       BA FRAGMENT.
FT   CHAIN      257    761       BB FRAGMENT.
FT   DOMAIN      33     96       SUSHI 1.
FT   DOMAIN      99    156       SUSHI 2.
FT   DOMAIN     161    216       SUSHI 3.
FT   DOMAIN     267    466       VWFA.
FT   DOMAIN     479    761       SERINE PROTEASE.
FT   ACT_SITE   523    523       CHARGE RELAY SYSTEM.
FT   ACT_SITE   573    573       CHARGE RELAY SYSTEM.
FT   ACT_SITE   696    696       CHARGE RELAY SYSTEM.
FT   DISULFID    34     73       BY SIMILARITY.
FT   DISULFID    59     95       BY SIMILARITY.
FT   DISULFID   100    142       BY SIMILARITY.
FT   DISULFID   128    155       BY SIMILARITY.
FT   DISULFID   162    202       BY SIMILARITY.
FT   DISULFID   188    215       BY SIMILARITY.
FT   CARBOHYD   119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   282    282       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   730    730       L -> P (IN REF. 4).
FT   CONFLICT   745    745       V -> E (IN REF. 4).
SQ   SEQUENCE   761 AA;  85604 MW;  9E661D44BA61020E CRC64;

  Query Match             17.4%;  Score 130.5;  DB 1;  Length 761;
  Best Local Similarity   28.1%;  Pred. No. 1.5e-05;
  Matches   38;  Conservative   18;  Mismatches   62;  Indels   17;  Gaps    6;

Qy       2 ISCDPPPEVKNA----RKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           | |  | : :|     | |:|:|       : : |   | | |     |    | :  ||
Db      98 IRCPRPQDFENGEFWPRSPFYNL- --SDQISFQCYDGYVLRGSANRIC    QENGRWD 149

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAIFRHCDSVIFLCKANFIMRGSKIVWCQAN 117
             ||:  :    | :| :| |    ||:  :|  | | ||: |      :::||:   ||
Db     150 GQTAICD--DGAGYCPNPGIPIGTRKVGSQ--YRLEDIVTYHCSRGLVLRGSQKRKCQEG 205

Qy     118 EMWGPTALPVCESDF 132
           |  |  | | :  |
Db     206 GSWSGTE-PSCQDSF 219


RESULT 32
LFC_FACTR
ID   LFC_FACTR       STANDARD;       PRT;  1019 AA.
AC   P28175;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Limulus clotting factor C precursor (EC 3.4.21.84) (FC).
OS   Tachypleus tridentatus (Japanese horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Tachypleus.
OX   NCBI_TaxID=6853;
```

```
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91177916; PubMed=2007602;
RA   Muta T., Miyata T., Misumi Y., Tokunaga F., Nakamura T., Toh Y.,
RA   Ikehara Y., Iwanaga S.;
RT   "Limulus factor C. An endotoxin sensitive serine protease zymogen
RT   with a mosaic structure of complement-like, epidermal growth
RT   factor like, and lectin-like domains.";
RL   J. Biol. Chem. 266:6554-6561(1991).
CC    -!- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC        SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC        ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC        FORM CATALYZES THE ACTIVATION OF FACTOR B.
CC    -!- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg-|-Ser-104 and
CC        124 Ile-|-Ile 125 bonds in Limulus clotting factor B to form
CC        activated factor B. Cleavage of Pro-Arg-| Xaa bonds in synthetic
CC        substrates.
CC    -!- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC        LIPOPOLYSACCHARIDES AND CHYMOTRYPSIN.
CC    -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC        BY A DISULFIDE BOND.
CC    -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC    -!- SIMILARITY: CONTAINS 1 C TYPE LECTIN FAMILY DOMAIN.
CC    -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC    -!- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC    -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC        TRYPSIN FAMILY.
CC   ---------- ... ----------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ... ----------------------------------------
DR   EMBL; D90271; BAA14315.1;  -.
DR   EMBL; D90272; BAA14316.1;  -.
DR   PIR; A38738; A38738.
DR   PIR; B38738; B38738.
DR   HSSP; P00734; 1UVU.
DR   MEROPS; S01.239; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000436; LCCL.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50820; LCCL; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Signal; Alternative splicing; Lectin;
KW   Hemolymph clotting; Glycoprotein; Cell adhesion; EGF-like domain;
KW   Sushi; Repeat.
FT   SIGNAL       1     25
FT   CHAIN       26   1019       LIMULUS CLOTTING FACTOR C.
FT   CHAIN       26    690       LIMULUS CLOTTING FACTOR C, HEAVY CHAIN.
FT   CHAIN      691   1019       LIMULUS CLOTTING FACTOR C, LIGHT CHAIN.
FT   CHAIN      691    762       LIMULUS CLOTTING FACTOR C, A CHAIN.
FT   CHAIN      763   1019       LIMULUS CLOTTING FACTOR C, B CHAIN.
FT   DOMAIN      102    137       EGF-LIKE.
FT   DOMAIN      142    195       SUSHI 1.
FT   DOMAIN      200    254       SUSHI 2.
FT   DOMAIN      260    321       SUSHI 3.
FT   DOMAIN      325    421       LCCL.
FT   DOMAIN      436    568       C-TYPE LECTIN.
FT   DOMAIN      576    634       SUSHI 4.
FT   DOMAIN      685    748       SUSHI 5.
FT   DOMAIN      763   1019       SERINE PROTEASE.
FT   ACT_SITE    809    809       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    865    865       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    966    966       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   BINDING     960    960       SUBSTRATE (BY SIMILARITY).
FT   DOMAIN      643    689       PRO-RICH.
FT   DISULFID    106    118       BY SIMILARITY.
FT   DISULFID    122    125       BY SIMILARITY.
FT   DISULFID    127    136       BY SIMILARITY.
FT   DISULFID    436    447       BY SIMILARITY.
FT   DISULFID    464    564       BY SIMILARITY.
FT   DISULFID    548    556       BY SIMILARITY.
FT   DISULFID    794    810       BY SIMILARITY.
FT   DISULFID    942    951       BY SIMILARITY.
FT   DISULFID    962    996       BY SIMILARITY.
FT   CARBOHYD    523    523       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    740    740       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    757    767       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    912    912       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    492    498       LTTTWIS -> TDNVTAT (IN SHORT ISOFORM).
FT   VARSPLIC    499   1019       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1019 AA;  112346 MW;  5B72B04C6715289B CRC64;

  Query Match             17.4%;  Score 130.5;  DB 1;  Length 1019;
  Best Local Similarity   26.6%;  Pred. No. 2.1e-05;
  Matches   42;  Conservative   22;  Mismatches   57;  Indels   37;  Gaps    9;

Qy       2 ISCDP-------PP----EVKNARKFYY- -    SLPIVIGIVLLYICSVSYRLIGEKAI 44
           ||| |       ||    |      | :    |  :  |  |||:|   | |||:  :
Db     180 ISCLPNGQWSSFPPKCIRECAKVSSPEHGEVNA-SGNM.EGAILRFSCDSPYYLIGQETL 239

Qy      45 FCISENQVHAIWDKAPPICESVNKTISCS-DF----  VPGGFMNKGSKAPFRHGDSV 96
           |    | :  |    | |:   | : |              | |  |  :  |   |  :
Db     240 TC----QGNGQWSGQTPQCK---KLVFCPELDP-NHAFHGVKICVEQKYGQFP--QGLEV 290

Qy      97 TFTCKANFTMKGSKTVWCQANEMWG---PTALPVCESD 131
           |:||  |: : |  |: |  :  |     |: :   : :
Db     291 TYTCSGNYFLMGFNTLKCNPDGSWSGSQPS-VECVADRE 328


RESULT 33
DAF_PONPY
ID   DAF_PONPY       STANDARD;       PRT;   440 AA.
AC   P49457;
DT   01-FEB-1998 (Rel. 33, Created)
DT   01-FEB-1998 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Complement decay-accelerating factor (CD55) (Fragment).
GN   DAF OR CD55.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94110622; PubMed=7506733;
RA   Nickells M.W., Alvarez J.I., Lublin D.M., Atkinson J.P.;
RT   "Characterization of DAF-2, a high molecular weight form of decay
RT   accelerating factor (DAF; CD55), as a covalently cross linked dimer
RT   of DAF-1.";
RL   J. Immunol. 152:676-685(1994).
CC   -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC       CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
```

```
CC       C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC       INTERACTION OF DAF WITH CELL ASSOCIATED C4B AND C3B POLYPEPTIDES
CC       INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND
CC       FACTOR B TO ENZYMATICALLY ACTIVE C2A AND B0 AND THEREBY PREVENTS
CC       THE FORMATION OF C4B2A AND C3BBB, THE AMPLIFICATION CONVERTASES OF
CC       THE COMPLEMENT CASCADE (BY SIMILARITY).
CC    -!- SUBUNIT: MONOMER (MAJOR FORM) AND NON DISULFIDE-LINKED, COVALENT
CC       HOMODIMER (MINOR FORM).
CC    -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI anchor.
CC    -!- ALTERNATIVE PRODUCTS: TWO FORMS OF DAF (DAF 2, SHOWN HERE, AND
CC       DAF 1) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC    -!- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR
CC       FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE
CC       ACTIVE SITE ON SCR3 (BY SIMILARITY).
CC    -!- PTM: THE SER/THR RICH DOMAIN IS HEAVILY O GLYCOSYLATED.
CC    -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC    -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62775; AAC60009.1; -.
DR   HSSP; P08604; 1HFI.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW   Alternative splicing; GPI anchor; Sushi.
FT   NON_TER       1      1
FT   CHAIN        -1    312       COMPLEMENT DECAY ACCELERATING FACTOR.
FT   PROPEP      313    340       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       -1     54       SUSHI 1.
FT   DOMAIN       56    118       SUSHI 2.
FT   DOMAIN      121    180       SUSHI 3.
FT   DOMAIN      183    243       SUSHI 4.
FT   DOMAIN      246    315       SER/THR RICH.
FT   DISULFID     34     53       BY SIMILARITY.
FT   DISULFID     57    104       BY SIMILARITY.
FT   DISULFID     88    117       BY SIMILARITY.
FT   DISULFID    122    163       BY SIMILARITY.
FT   DISULFID    149    179       BY SIMILARITY.
FT   DISULFID    184    226       BY SIMILARITY.
FT   DISULFID    212    242       BY SIMILARITY.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    107    107       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       312    312       GPI-ANCHOR (BY SIMILARITY).
SQ   SEQUENCE   340 AA;  37180 MW;  D4D8665C058204290 CRC64;

  Query Match          17.8%;  Score 130;  DB 1;  Length 340;
  Best Local Similarity   28.0%;  Pred. No. 7e-06;
  Matches   47;  Conservative   25;  Mismatches   94;  Indels   16;  Gaps    8;

Qy       3 SCLPPTEVKNARKEYYSGETIVLECVEGYTLIGEAVIRCESNGQVGAWSGAFP 61
           ||  | |:|:|  :    |  |:|  : :::|:  |:|| |  : | ||  :|    |
Db     121 SCPNPGETRNGQKD-VSNGTLFGAILSFSCNTGYKLFGPTSSLCLISGSSVQ  WSDPLP 177

Qy      62 ICESVNKIISS SDPIVGKVNFMNRGSKAPPPHGDSVTFTCKANPTMKGSKTVWCQAN    117
           |    : | | : |  :   | : :| :  : :  |:|: |   :|| |  :::| |
Db     178 EC -  REIYCPAPPQIDNGII QGKRDHYGYRQSITYACNKGYTMIGEHSIYCTVNDDE 232

Qy     118  EMWGPTALPVC 128
            :  ||   | |
Db     233 GEWSGPP  PEC 242


RESULT 34
TEM5_RAT
```

```
ID   TEM5_RAT        STANDARD;      PRT:   768 AA.
AC   P98106;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   P-selectin precursor (Granule membrane protein 140) (GMP 140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   SELP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=94333817; PubMed=7520013;
RA   Auchampach J.A., Oliver M.G., Anderson D.C., Manning A.M.;
RT   "Cloning, sequence comparison and in vivo expression of the gene
RT   encoding rat P-selectin.";
RL   Gene 145:251-255(1994).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED: SPLEEN,
CC       LUNG, BRAIN, LIVER, HEART, KIDNEY, THYMUS, SMALL INTESTINE.
CC   -!- INDUCTION: BY ACUTE INFLAMATION (PROBABLE).
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS; RAT P-LECTIN LACKS THE
CC       HUMAN SUSHI-2 EQUIVALENT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L23088; AAA60325.1; -.
DR   HSSP; P16109; 1FSB.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00443; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat; Lipoprotein; Palmitate.
FT   SIGNAL        1     41       POTENTIAL.
FT   CHAIN        42    768       P-SELECTIN.
FT   DOMAIN       42    709       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    710    733       POTENTIAL.
FT   DOMAIN       734    768       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN       159    195       EGF-LIKE.
FT   DOMAIN       199    258       SUSHI 1.
FT   DOMAIN       261    320       SUSHI 2.
FT   DOMAIN       323    382       SUSHI 3.
FT   DOMAIN       385    444       SUSHI 4.
FT   DOMAIN       447    506       SUSHI 5.
FT   DOMAIN       509    568       SUSHI 6.
```

```
FT   DOMAIN       579    638       SUSHI 7.
FT   DOMAIN       641    700       SUSHI 8.
FT   DISULFID      60    158       BY SIMILARITY.
FT   DISULFID     131    150       BY SIMILARITY.
FT   DISULFID     168    183       BY SIMILARITY.
FT   DISULFID     185    194       BY SIMILARITY.
FT   DISULFID     200    244       BY SIMILARITY.
FT   DISULFID     230    257       BY SIMILARITY.
FT   DISULFID     262    306       BY SIMILARITY.
FT   DISULFID     292    319       BY SIMILARITY.
FT   DISULFID     324    368       BY SIMILARITY.
FT   DISULFID     354    381       BY SIMILARITY.
FT   DISULFID     386    430       BY SIMILARITY.
FT   DISULFID     416    443       BY SIMILARITY
FT   DISULFID     448    492       BY SIMILARITY.
FT   DISULFID     478    505       BY SIMILARITY.
FT   DISULFID     510    554       BY SIMILARITY.
FT   DISULFID     540    567       BY SIMILARITY.
FT   DISULFID     580    624       BY SIMILARITY.
FT   DISULFID     610    637       BY SIMILARITY.
FT   DISULFID     642    686       BY SIMILARITY.
FT   DISULFID     672    699       BY SIMILARITY.
FT   CARBOHYD      45     45       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     107    107       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     347    347       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     654    654       N-LINKED (GLCNAC.     ) (POTENTIAL)
FT   CARBOHYD     661    661       N-LINKED (GLCNAC      ) (POTENTIAL)
FT   CARBOHYD     679    679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID        745    745       PALMITATE (BY SIMILARITY).
FT   SITE         756    759       ENDOCYTOSIS SIGNAL (PROBABLE).
SQ   SEQUENCE   768 AA;  83517 MW;  26FD7E8A5F3F1316 CRC64;

  Query Match          17.2%;  Score 129.5;  DB 1;  Length 768;
  Best Local Similarity   27.4%;  Pred. No. 1.9e-05;
  Matches   32;  Conservative   23;  Mismatches   41;  Indels   21;  Gaps    7;

Qy      25 GTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPPICESVNK----TISCSDPIVPGG 80
           |:   ::|:  : |:| : : |      |   |   || |: |       :: |     ||
Db     533 GSTCHFSCNEEFELLGSRNVEC----IVSGRWSAPPPICKGVTSLPVPSVRCPALTPG- 587

Qy      81 FMNKGSKAPFRH----GDSVT  FTCKANFTMKGSKTVWCQANEMWGPTAL-PVCES 130
             :|: :   |    | : |  | ||  ||:::|: :: | |:  |  ||: ||| :
Db     588 -- QGTMSCRHHLESFGFNTICYFGCKIGFILRGANSLRCGASGQW- TAVTPVCRA 639


RESULT 35
LEM2_CANFA
ID   LEM2_CANFA     STANDARD;      PRT;   611 AA.
AC   P33730;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1996 (Rel. 34, Last annotation update)
DE   E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Jugular vein;
RA   Manning A.M., Lane C.L., Auchampach J.A., Kukielka G.L.,
RA   Rosenbloom C.L., Anderson D.C.;
RL   Submitted (XXX-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1-3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L23087, AAA30843.1; -.
DR   HSSP; P16581, 1ESL.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00463; SELECTIN.
DR   SMART; SM00032; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL       1     22       BY SIMILARITY
FT   CHAIN       23    611       E-SELECTIN.
FT   DOMAIN      23    557       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   558    579       POTENTIAL.
FT   DOMAIN     580    611       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      39    139       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN     140    176       EGF-LIKE
FT   DOMAIN     180    239       SUSHI 1
FT   DOMAIN     242    301       SUSHI 2
FT   DOMAIN     304    364       SUSHI 3
FT   DOMAIN     367    427       SUSHI 4
FT   DOMAIN     430    490       SUSHI 5
FT   DOMAIN     493    549       SUSHI 6
FT   DISULFID    41    139       BY SIMILARITY.
FT   DISULFID   112    131       BY SIMILARITY.
FT   DISULFID   144    155       BY SIMILARITY.
FT   DISULFID   149    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   181    225       BY SIMILARITY.
FT   DISULFID   211    238       BY SIMILARITY.
FT   DISULFID   243    287       BY SIMILARITY.
FT   DISULFID   273    300       BY SIMILARITY.
FT   DISULFID   305    350       BY SIMILARITY.
FT   DISULFID   336    363       BY SIMILARITY.
FT   DISULFID   368    413       BY SIMILARITY.
FT   DISULFID   399    426       BY SIMILARITY.
FT   DISULFID   431    476       BY SIMILARITY.
FT   DISULFID   462    489       BY SIMILARITY.
FT   DISULFID   494    535       BY SIMILARITY.
FT   DISULFID   521    548       BY SIMILARITY.
FT   CARBOHYD    26     26       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   204    204       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   266    266       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   313    313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   611 AA;  66314 MW;  35DA9E3DF225E4F6 CRC64;
```

```
  Query Match             [illegible]%;  Score 129;  DB 1;  Length 611;
  Best Local Similarity   29.1%;  Pred. No. 1.7e-05;
  Matches   80;  Conservative   22;  Mismatches   41;  Indels   10;  Gaps    5;

Qy      60 YTCSESYKLIGEKAIPCTSENQVHAIWDKAPPICESVNKTISCSDPIVPGSEMN  KGSK 87
           :|||: : : |  | | | :|  |  | :  |:|:: ::  : | |    |:|:  :|:
Db     434 FTCNEGFLMQGPAQIECTAQGQ    WSQQVPVCKAVSCPKALSSP  ERQYMSCLPGAS 486

Qy      88 APERHGDSVTFTCKANFTMKGSKTVWCQANEMWGPTALPVCES 130
           | | | | |:  : ::::: :     |  :  | ||:
Db     487 GSPQSGSSCTEFCFRGIVLNGSKTLQWLTGKW DSEEPTCEA 528


RESULT 36
LEM3_SHEEP
ID   LEM3_SHEEP      STANDARD;      PRT:   769 AA.
AC   P08109;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   E-selectin precursor (Granule membrane protein 140) (GMP 140) (PADGEM)
DE   (CD62E) (Leukocyte endothelial cell adhesion molecule 2) (LECAM2).
GN   SELE.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RA   Burns S.A., Neuteld E.L., Donady J.J.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CA(2+) DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L34277; AAB59261.1; -.
DR   HSSP; P16109; 1FSD.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00448; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     32       POTENTIAL.
FT   CHAIN        33    769       E-SELECTIN.
FT   DOMAIN       33    717       EXTRACELLULAR (POTENTIAL)
FT   TRANSMEM    718    734       POTENTIAL.
FT   DOMAIN      735    769       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58    158       C-TYPE LECTIN.
FT   DOMAIN      159    195       EGF-LIKE.
FT   DOMAIN      199    258       SUSHI 1.
FT   DOMAIN      261    320       SUSHI 2.
FT   DOMAIN      323    382       SUSHI 3.
FT   DOMAIN      385    444       SUSHI 4.
FT   DOMAIN      447    506       SUSHI 5.
FT   DOMAIN      509    568       SUSHI 6.
FT   DOMAIN      580    639       SUSHI 7.
FT   DOMAIN      642    701       SUSHI 8.
FT   DISULFID     60    158       BY SIMILARITY.
FT   DISULFID    131    150       BY SIMILARITY.
FT   DISULFID    163    174       BY SIMILARITY.
FT   DISULFID    168    183       BY SIMILARITY.
FT   DISULFID    185    194       BY SIMILARITY.
FT   DISULFID    200    244       BY SIMILARITY.
FT   DISULFID    230    257       BY SIMILARITY.
FT   DISULFID    262    306       BY SIMILARITY.
FT   DISULFID    292    319       BY SIMILARITY.
FT   DISULFID    324    368       BY SIMILARITY.
FT   DISULFID    354    381       BY SIMILARITY.
FT   DISULFID    386    430       BY SIMILARITY.
FT   DISULFID    416    443       BY SIMILARITY.
FT   DISULFID    448    492       BY SIMILARITY.
FT   DISULFID    478    505       BY SIMILARITY.
FT   DISULFID    510    554       BY SIMILARITY.
FT   DISULFID    540    567       BY SIMILARITY.
FT   DISULFID    581    625       BY SIMILARITY.
FT   DISULFID    611    638       BY SIMILARITY.
FT   DISULFID    643    687       BY SIMILARITY.
FT   DISULFID    673    700       BY SIMILARITY.
FT   CARBOHYD     54     54       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    219    219       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    347    347       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    604    604       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    655    655       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    662    662       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE        757    760       ENDOCYTOSIS SIGNAL (PROBABLE).
FT   VARIANT     566    566       S -> I.
FT   VARIANT     579    579       I -> V.
SQ   SEQUENCE   769 AA;  84817 MW;  24E4257D60FAB1F0 CRC64;

  Query Match             17.1%;  Score 128.5;  DB 1;  Length 769;
  Best Local Similarity   26.2%;  Pred. No. 2.4e-05;
  Matches   39;  Conservative   18;  Mismatches   59;  Indels   33;  Gaps    8;

Qy       6 PPTEVK    NAEEPYYCLP  IVP                 ATVLSYTVCEQYSQLCHEEA 43
           |||  |    :|  |   |  | ||                 |   : :   : |:|| |
Db     563 PPPSCKVDIVSAPAPGLRCVSLIAPNQGIMSCQHHGRNFGLNFTCHECGCKAGFTMGESA 622

Qy      44 IFCTSENQVRAIWGKAEPICESVN RISCSGHVSQGRMNKCSA AVEHQGSVIECK 101
           : |    |   |   | | :|   :  ::|||   ||      | | : :| | |
Db     623 LQCKPSRQ   -WIAVAPICRAVKCSKLEVLEPIV    MNCSNEWGNESSLSICSFIPCP 674

Qy     102 ANFTMKGSKIVWCQANEMWGPTALPVCES 130
           : ||:  || |  |  | : : |::
Db     675 EGQLCNGSERTACQENGQWS-ITMPTCQA 702


RESULT 37
LEC_CARRO
ID   LEC_CARRO       STANDARD;       PRT:  1019 AA.
AC   Q26422;
DT   30-MAY-2000 (Rel. 49, Created)
DT   30-MAY-2000 (Rel. 49, Last sequence update)
```

```
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Limulus clotting factor C precursor (EC 3.4.21.84) (FC).
OS   Carcinoscorpius rotundicauda (Southeast Asian horseshoe crab).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC   Limulidae; Carcinoscorpius.
OX   NCBI TaxID=6848;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=95268506; PubMed=7538401;
RA   Ding J.L., Navas M.A. III, Ho B.;
RT   "Molecular cloning and sequence analysis of factor C cDNA from the
RT   Singapore horseshoe crab, Carcinoscorpius rotundicauda.";
RL   Mol. Mar. Biol. Biotechnol. 4:90-103(1995).
CC   -!- FUNCTION: THIS ENZYME IS CLOSELY ASSOCIATED WITH AN ENDOTOXIN-
CC       SENSITIVE HEMOLYMPH COAGULATION SYSTEM WHICH MAY PLAY IMPORTANT
CC       ROLES IN BOTH HEMOSTASIS AND HOST DEFENSE MECHANISMS. ITS ACTIVE
CC       FORM CATALYZES THE ACTIVATION OF FACTOR B
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of 103-Arg-|-Ser-104 and
CC       124 Ile-|-Ile 125 bonds in Limulus clotting factor B to form
CC       activated factor B. Cleavage of Pro-Arg-|-Xaa bonds in synthetic
CC       substrates.
CC   -!- ENZYME REGULATION: ACTIVATED BY GRAM-NEGATIVE BACTERIAL
CC       LIPOPOLYSACCHARIDES AND CHYMOTRYPSIN (BY SIMILARITY).
CC   -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC       BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 LCCL DOMAIN.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------- ---------- ----------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ---------- ----------
DR   EMBL; S77064; AAB34361.1; -.
DR   HSSP; P00744; 1UVU.
DR   MEROPS; S01.219; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000561; EGF-like
DR   InterPro; IPR000043; LCCL.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 5.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00032; CCP; 5.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50820; LCCL; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Signal; Lectin; Hemolymph clotting;
KW   Glycoprotein; Cell adhesion; EGF-like domain; Sushi; Repeat.
FT   SIGNAL       1     25       BY SIMILARITY.
FT   CHAIN       26   1019       LIMULUS CLOTTING FACTOR C.
FT   CHAIN       26    690       LIMULUS CLOTTING FACTOR C, HEAVY CHAIN.
FT   CHAIN      691   1019       LIMULUS CLOTTING FACTOR C, LIGHT CHAIN.
FT   CHAIN      691    762       LIMULUS CLOTTING FACTOR C, A CHAIN.
FT   CHAIN      763   1019       LIMULUS CLOTTING FACTOR C, B CHAIN.
FT   DOMAIN     102    137       EGF-LIKE.
FT   DOMAIN     142    195       SUSHI 1.
FT   DOMAIN     200    254       SUSHI 2.
FT   DOMAIN     260    321       SUSHI 3.
FT   DOMAIN     325    421       LCCL.
FT   DOMAIN     436    568       C-TYPE LECTIN.
FT   DOMAIN     576    634       SUSHI 4.
FT   DOMAIN     685    748       SUSHI 5.
FT   DOMAIN     763   1019       SERINE PROTEASE.
FT   ACT_SITE   809    809       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   865    865       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   966    966       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   BINDING    960    960       SUBSTRATE (BY SIMILARITY).
FT   DOMAIN     643    689       PRO-RICH.
FT   DISULFID   106    118       BY SIMILARITY.
FT   DISULFID   112    125       BY SIMILARITY.
FT   DISULFID   127    136       BY SIMILARITY.
FT   DISULFID   436    447       BY SIMILARITY.
FT   DISULFID   464    564       BY SIMILARITY.
FT   DISULFID   538    556       BY SIMILARITY.
FT   DISULFID   794    810       BY SIMILARITY.
FT   DISULFID   932    951       BY SIMILARITY.
FT   DISULFID   962    996       BY SIMILARITY.
FT   CARBOHYD   523    523       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   740    740       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   767    767       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   912    912       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1019 AA;  112429 MW;  9187ED8B817B6C3 CRC64;

  Query Match             17.0%;  Score 72.5;  DB 1;  Length 1019;
  Best Local Similarity   24.7%;  Pred. No. 4.1e-05;
  Matches   39;  Conservative   28;  Mismatches   54;  Indels   47;  Gaps    8;

Qy       2 ISCDP-------PPEV----------KNARKPVYSLPIVPGTVLRYTCSPSYRLIGEKAI 44
           ||| |       ||:           : :    |   : |  ||:||   | |||: :
Db     180 ISCLPNGQWSNFPPKCIRECAMVSSPEHGKVNALSCDMIECAIIRFSCDSPYYLIGQEII 239

Qy      45 FCISHNQVHATWDKAPPICESVNKTISCS--DPI---  -VPGGFMNKGSKAPFRHGDSV 96
             |    | : |:   | |:::   : |    :      | |   |  : |   :  :
Db     240 TC----QGNGQWNGQIPQCKNL---VFCPDLDP-VNHAEHKVKIGVEQKYGQFP--QGTEV 290

Qy      97 TFTCKANFTMKGSKTVWCQANEMWG---PTALPVCESD 131
           |:|| |: : |  |: | |  : |     |:   | : :
Db     291 TYTCSGNYFLMGFDTLKCNPDGSWSGSQPSCVKVADKE 328


RESULT 38
APOH_BOVIN
ID   APOH_BOVIN     STANDARD;      PRT;   345 AA
AC   P17690; Q28052;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI).
GN   APOH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Gao B., Virmani M., Pomm E., Lazar-Wesley E., Sakaguchi K.,
RA   Appella E., Kunos G., Takacs L.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 4-345 FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RC   TISSUE=Liver;
```

```
RX   MEDLINE; 92242647; PubMed; 1567819;
RA   Bendixen E., Halkier T., Magnusson S., Sottrup-Jensen L.,
RA   Kristensen T.;
RT   "Complete primary structure of bovine beta 2-glycoprotein 1:
RT   localization of the disulfide bridges.";
RL   Biochemistry 31:3611-3617(1992).
RN   [3]
RP   SEQUENCE OF 20-345, CARBOHYDRATE LINKAGE SITES, AND DISULFIDE BONDS.
RC   TISSUE=Plasma;
RX   MEDLINE=92089075; PubMed=1751487;
RA   Kato H., Enjyoji K.-I.;
RT   "Amino acid sequence and location of the disulfide bonds in bovine
RT   beta 2-glycoprotein I: the presence of five Sushi domains.";
RL   Biochemistry 30:1168-1174(1991).
RN   [4]
RP   SEQUENCE OF 20-41.
RX   MEDLINE=90226428; PubMed=2327984;
RA   Lu Q., Boucher R., Esch F., Compote L.F.;
RT   "Isolation from fetal bovine serum of an apolipoprotein H-like
RT   protein which inhibits thymidine incorporation in fetal calf
RT   erythroid cells.";
RL   Biochem. J. 267:261-264(1990).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07803; AAA30822.1; -.
DR   EMBL; X60065; CAA42669.1; -.
DR   PIR; S23567; NBBO.
DR   PIR; S09042; S09042.
DR   HSSP; P02749; 1CLZ.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19
FT   CHAIN       20    345       BETA-2 GLYCOPROTEIN I.
FT   DOMAIN      22     80       SUSHI 1.
FT   DOMAIN      83    138       SUSHI 2.
FT   DOMAIN     141    201       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     263    345       SUSHI-LIKE.
FT   DISULFID    23     66
FT   DISULFID    51     74
FT   DISULFID    84    124
FT   DISULFID   110    137
FT   DISULFID   142    188
FT   DISULFID   174    200
FT   DISULFID   205    248
FT   DISULFID   234    260
FT   DISULFID   264    315
FT   DISULFID   300    325
FT   DISULFID   307    345
FT   CARBOHYD    92     92       N-LINKED (GLCNAC...).
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...).
FT   CARBOHYD   183    183       N-LINKED (GLCNAC...).
FT   CARBOHYD   193    193       N-LINKED (GLCNAC...).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC...).
FT   CONFLICT   101    101       F -> G (IN REF. 1).
FT   CONFLICT   108    108       F -> S (IN REF. 1).
FT   CONFLICT   177    177       H -> R (IN REF. 1).
FT   CONFLICT   191    191       H -> N (IN REF. 3).
FT   CONFLICT   194    194       W -> C (IN REF. 1).
FT   CONFLICT   259    259       S -> N (IN REF. 1).
FT   CONFLICT   302    302       H -> N (IN REF. 1).
FT   CONFLICT   305    305       K -> R (IN REF. 1).
FT   CONFLICT   329    329       H -> R (IN REF. 1).
SQ   SEQUENCE   345 AA;  38252 MW;  E1172DAB60946C33 CRC64;

  Query Match           16.9%;  Score 127;  DB 1;  Length 345;
  Best Local Similarity  25.9%;  Pred. No. 1.4e-05;
  Matches   56;  Conservative   20;  Mismatches   65;  Indels   18;  Gaps    6;

Qy       2 LSCPPPEVKNAFKDPYYSLEYIVF   - CIVLEYDCSPSYRE CEKAIFCTSENQVHATW 56
           |:| |||  | |     |  |:      |:   : : | : :|   : :       | |
Db     140 LTCPPPTPEEASCSVYK-LACNNSFYC-KAVES-QGVMAMEGNEVE-D     HGNW 194

Qy      57 DEAryCESVNEIDCSCPIVP-GPEMNPCPKAFNKDCYVLI KANIMKCCKIVWCQ 115
           : |   |    : |  | |  |||:  :     : : |||  |   ::: | : | |
Db     195 TQLPECRE-   VRCPFPSRPDNGFVNHDANPVLYYKDTATFGCHETYSLDGPEEVECS 249

Qy     116 ANEMWGPTALPVCESDFPL 134
                :  :| | |::   |
Db     250 KFGNW  SAQPSCKASCKL 266


RESULT 39
FHR1_HUMAN
ID   FHR1_HUMAN     STANDARD;      PRT:   330 AA.
AC   Q03591; Q9UJ17;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement factor H-related protein 1 precursor (FHR-1) (H factor like
DE   protein 1) (H-factor like 1) (H36).
GN   FHR1 OR HFL1 OR CFHL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91201892; PubMed=1826708;
RA   Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;
RT   "Cloning of the 1.4-kb mRNA species of human complement factor H
RT   reveals a novel member of the short consensus repeat family related
RT   to the carboxy terminal of the classical 150 kDa molecule.";
RL   J. Immunol. 146:3190-3196(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20245597; PubMed=10781884;
RA   Male D.A., Ormsby R.J., Ranganathan S., Giannakis E., Gordon D.L.;
RT   "Complement factor H: sequence analysis of 221 kb of human genomic DNA
RT   containing the entire FH, FHR-1 and FHR-4 genes.";
RL   Mol. Immunol. 37:41-52(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hall R.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 4-330 FROM N.A.
RX   MEDLINE=91268081; PubMed=1711047;
RA   Skerka C., Horstmann R.D., Zipfel P.F.;
RT   "Molecular cloning of a human serum protein structurally related to
RT   complement factor H.";
RL   J. Biol. Chem. 266:12015-12020(1991).
RN   [5]
RP   CARBOHYDRATES.
RX   MEDLINE=91123700; PubMed=1825108;
RA   Timmann C., Leippe M., Horstmann R.D.;
RT   "Two major serum components antigenically related to complement factor
RT   H are different glycosylation forms of a single protein with no
RT   factor H-like complement regulatory functions.";
RL   J. Immunol. 146:1265-1270(1991).
```

```
RN   [6]
RP   REVIEW.
RX   MEDLINE=94226679; PubMed 8172644;
RA   Ziptel P.F., Skerka C.;
RT   "Complement factor H and related proteins: an expanding family of
RT   complement-regulatory proteins?";
RL   Immunol. Today 15:121 126(1994).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN
CC       ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID
CC       METABOLISM.
CC    !  SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: LIVER.
CC   -!- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE
CC       SIDE CHAIN AND THE OTHER WITH TWO.
CC    !  SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC    !  SIMILARITY: STRONG, TO FACTOR H.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M65292; AAA35946.1; -.
DR   EMBL; M65293; AAA35947.1; -.
DR   EMBL; AL049741; CAB53063.1; -.
DR   EMBL; X56209; CAA39666.1; -.
DR   PIR; A40455; A40455.
DR   PIR; S14604; S14604.
DR   HSSP; P10998; 1VVD.
DR   MIM; 134371; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
KW   Repeat; Glycoprotein; Sushi; Signal; Polymorphism.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    330       COMPLEMENT FACTOR H-RELATED PROTEIN 1.
FT   DOMAIN       22     84       SUSHI 1.
FT   DOMAIN       86    141       SUSHI 2.
FT   DOMAIN      146    202       SUSHI 3.
FT   DOMAIN      207    264       SUSHI 4.
FT   DOMAIN      265    328       SUSHI 5.
FT   CARBOHYD    126    126       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .).
FT   VARIANT     157    157       Y -> H.
FT                                /FTId=VAR_001980.
FT   VARIANT     159    159       V -> L.
FT                                /FTId=VAR_001981.
FT   VARIANT     175    175       Q -> E.
FT                                /FTId=VAR_001982.
FT   CONFLICT     71     71       T -> N (IN REF. 4).
SQ   SEQUENCE   330 AA;  37661 MW;  8DC0D3F92A85E035 CRC64;

  Query Match          16.7%;  Score 125.5;  DB 1;  Length 330;
  Best Local Similarity   26.5%;  Pred. No. 1.9e-05;
  Matches   85;  Conservative   25;  Mismatches   59;  Indels   13;  Gaps    7;

Qy       1 EISCDPPPEVKNARKPYYSLPIVP GTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKA 59
           : ||  || |:||      :  | |  :|| |   | : |:: : |:: |     | :
Db     144 DTSCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGN-----WTE- 197

Qy      60 PPICESVNKTISCS DPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANE 118
             || |:   . |  |   |  :  | :      : :    || : |:  : ::|:| : |: |
Db     198 PPQCK- DSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCR-NG 254

Qy     119 MWG  PTALPVC 128
             |   |  |  |
Db     255 QWSEPPKCLHPC 266


RESULT 40
```

```
LEM3_BOVIN
ID   LEM3_BOVIN     STANDARD;      PRT;   646 AA.
AC   P42201;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   P selectin precursor (Granule membrane protein 140) (GMP-140) (PADGEM)
DE   (CD62P) (Leukocyte-endothelial cell adhesion molecule 3) (LECAM3).
GN   S-LP.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Capillary endothelium;
RX   MEDLINE=93249394; PubMed=7683458;
RA   Strubel N.A., Nguyen M., Kansas G.S., Tedder T.F., Bischoff J.;
RT   "Isolation and characterization of a bovine cDNA encoding a
RT   functional homolog of human P-selectin.";
RL   Biochem. Biophys. Res. Commun. 192:338-344(1993).
CC   -!- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC       TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC       INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC       LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC       AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC       ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC       THE CELL SURFACE.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECTINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS; BOVINE P-LECTIN LACKS
CC       THE HUMAN SUSHI-3, -4 AND -7 EQUIVALENTS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L12041; AAA30743.1; -.
DR   HSSP; P16109; 1FSB.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 6.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 6.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL        1     41       BY SIMILARITY
FT   CHAIN        42    646       P-SELECTIN.
FT   DOMAIN       42    587       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    588    611       POTENTIAL.
FT   DOMAIN      612    646       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       58    158       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      159    195       EGF-LIKE.
FT   DOMAIN      199    258       SUSHI 1.
FT   DOMAIN      261    320       SUSHI 2.
```

```
FT   DOMAIN        424     482        SUSHI 3.
FT   DOMAIN        485     444        SUSHI 4.
FT   DOMAIN        457     516        SUSHI 5.
FT   DOMAIN        529     578        SUSHI 6.
FT   DISULFID       60     158        BY SIMILARITY.
FT   DISULFID      131     150        BY SIMILARITY.
FT   DISULFID      164     174        BY SIMILARITY.
FT   DISULFID      168     183        BY SIMILARITY.
FT   DISULFID      185     194        BY SIMILARITY.
FT   DISULFID      200     244        BY SIMILARITY.
FT   DISULFID      230     257        BY SIMILARITY.
FT   DISULFID      262     306        BY SIMILARITY.
FT   DISULFID      292     319        BY SIMILARITY.
FT   DISULFID      324     368        BY SIMILARITY.
FT   DISULFID      354     381        BY SIMILARITY.
FT   DISULFID      386     430        BY SIMILARITY.
FT   DISULFID      416     443        BY SIMILARITY.
FT   DISULFID      468     562        BY SIMILARITY.
FT   DISULFID      488     515        BY SIMILARITY.
FT   DISULFID      520     564        BY SIMILARITY.
FT   DISULFID      550     527        BY SIMILARITY.
FT   CARBOHYD       48      48        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       54      54        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       80      80        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      180     180        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      212     212        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      219     219        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      336     336        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      481     481        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      542     542        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      549     549        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      557     557        N-LINKED (GLCNAC...) (POTENTIAL).
FT   SITE          634     637        ENDOCYTOSIS SIGNAL (PROBABLE).
SQ   SEQUENCE   646 AA;  71229 MW;  57A912A4627A6ACA CRC64;

  Query Match             16.5%;  Score 124.5;  DB 1;  Length 646;
  Best Local Similarity   28.0%;  Pred. No. 5e-05;
  Matches   84;  Conservative   18;  Mismatches   48;  Indels   11;  Gaps    5;

Qy       29 IVLRYTCSPSYRLIGEKAIFCISPNQVHAIWDKAPPICESVN KTISCSDPIVPGGFMNK 84
            |    |  : |:|:  : :    |   | | | | :|    :  :::||    ||
Db      482 FICHPGCKKAGFTLLGDSALQCRPSRQ    WTAAAPTCRAVKCAKLPVIEPIV  MNC 534

Qy       85 GSK APFRHGDSVIFTCKANFIMKGSKIVWCQANEMWGPTALPVCES 130
            :    | :| : :| |     : ||:   || | |  | : :| |::
Db      535 SNPWGNESYGSTCSFHCPEGQLLNGSERIVCQENGQWS EIMPTCQA 579


RESULT 4
APOH_MOUSE
ID   APOH_MOUSE      STANDARD;      PRT;   345 AA.
AC   Q01339;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI) (Activated protein C-binding protein) (APC inhibitor).
GN   APOH OR BCHI.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94372004; PubMed=8039487;
RA   Nomaka M., Matsuda Y., Shiroishi T., Moriwaki K., Natsuume-Sakai S.;
RT   "Molecular cloning of mouse beta-2-glycoprotein I and mapping of the
RT   gene to chromosome 11.";
RL   Genomics 19:1082-1087(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/J; TISSUE=Liver;
RX   MEDLINE=94242017; PubMed=7514402;
RA   Sellar G.C., Steel D.M., Zafiropoulos A., Seery L.T.,
RA   Whitehead A.S.;
RT   "Characterization, expression and evolution of mouse beta-2
RT   glycoprotein I (apolipoprotein H).";
RL   Biochem. Biophys. Res. Commun. 200:1521-1528(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Liver;
RA   Kristensen T.;
RT   "Structure of the human beta-2 glycoprotein I gene.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -----------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------
DR   EMBL; D10656; BAA00945.1; -.
DR   EMBL; S70439; AAB30789.1; -.
DR   EMBL; Y17356; CAA72190.1; -.
DR   PIR; A43286; NBMS.
DR   HSSP; P02749; 1QIZ.
DR   MGD; MGI:88058; Apoh.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL          1     19
FT   CHAIN          20    345        BETA-2-GLYCOPROTEIN I.
FT   DOMAIN         22     80        SUSHI 1.
FT   DOMAIN         83    138        SUSHI 2.
FT   DOMAIN        141    201        SUSHI 3.
FT   DOMAIN        204    261        SUSHI 4.
FT   DOMAIN        263    345        SUSHI-LIKE.
FT   DISULFID       23     66        BY SIMILARITY.
FT   DISULFID       51     79        BY SIMILARITY.
FT   DISULFID       84    124        BY SIMILARITY.
FT   DISULFID      110    137        BY SIMILARITY.
FT   DISULFID      142    188        BY SIMILARITY.
FT   DISULFID      174    200        BY SIMILARITY.
FT   DISULFID      205    248        BY SIMILARITY.
FT   DISULFID      234    260        BY SIMILARITY.
FT   DISULFID      264    315        BY SIMILARITY.
FT   DISULFID      300    325        BY SIMILARITY.
FT   DISULFID      307    345        BY SIMILARITY.
FT   CARBOHYD      162    162        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      183    183        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      193    193        N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT      252    252        G   R (IN REF. 2).
SQ   SEQUENCE   345 AA;  38619 MW;  CB4E8A6E8D51C940 CRC64;

  Query Match             16.5%;  Score 124;  DB 2;  Length 345;
  Best Local Similarity   24.5%;  Pred. No. 2.8e-05;
  Matches   84;  Conservative   21;  Mismatches   66;  Indels   38;  Gaps    6;

Qy        2 ISCDPPPE     VKNARKPYYSLEPIVPGEIVEKVICSPSYRLIGEKAIFCISPNQVHAIW 50
            |:| |||      :|: |   :  :   ||: : | | : :|   : | :     |
Db      140 ITCPPPEVEFAIEPDYPPSRAGNNSIYGPDVV FKCLPELIAMIGNDIVMKIEQ    GNW 194

Qy       51 DKAPPICESVNKTISCSDPIVPGGFMNKGSKAIFFRHGDSVIFICKANFIMKGSKIVWCQ 110
            : |  |      : |  | :| |::|  :|   :| | |  |  : |  |     |
Db      195 IRLPECLE     VKCPFPDRFENGYVNYIAKIVEIYKEKAEGCDHEIYKLDGGLEAECL 249

Qy      111 ANEMWGPTALPVCESDFPL 134
```

```
                |  : || |     |
Db    250 KTGTW--SFLPTCRESCKL 266


RESULT 42
LEM2_PIG
ID   LEM2_PIG        STANDARD;       PRT;   484 AA.
AC   P98110;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   E selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE   (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE   (CD62E).
GN   SELE.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aortic endothelium;
RX   MEDLINE=95071392; PubMed=7526854;
RA   Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RA   Matis L.A., Rother R.P.;
RT   "Molecular and functional analysis of porcine E-selectin reveals a
RT   potential role in xenograft rejection.";
RL   Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aortic endothelium;
RX   MEDLINE=94271236; PubMed=7516159;
RA   Tsang Y.T.M., Haskard D.O., Robinson M.K.;
RT   "Cloning and expression kinetics of porcine vascular cell adhesion
RT   molecule.";
RL   Biochem. Biophys. Res. Commun. 201:805-805(1994).
CC   -!- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC       MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC       ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC       POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC       GLYCOLIPIDS).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC       REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: TO OTHER SELECTINS/LECAMS.
CC   -!- SIMILARITY: CONTAINS 1 C TYPE LECTIN FAMILY DOMAIN
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS; PORCINE E-LECTIN LACKS
CC       THE HUMAN SUSHI-1 AND -4 EQUIVALENTS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L39076; AAA61545.1; -.
DR   EMBL; U08350; AAA21541.1; -.
DR   HSSP; P16581; 1ESL.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR001304; lectin_c.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 4.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 4.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW   Selectin; Signal; Sushi; Repeat.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23    484       E-SELECTIN.
FT   DOMAIN      23    429       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    430   451       POTENTIAL.
FT   DOMAIN      452   484       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39   139       C-TYPE LECTIN (SHORT FORM).
FT   DOMAIN      140   176       EGF-LIKE.
FT   DOMAIN      180   236       SUSHI 1.
FT   DOMAIN      239   299       SUSHI 2.
FT   DOMAIN      302   362       SUSHI 3.
FT   DOMAIN      365   421       SUSHI 4.
FT   DISULFID     41   139       BY SIMILARITY.
FT   DISULFID    112   131       BY SIMILARITY.
FT   DISULFID    144   155       BY SIMILARITY.
FT   DISULFID    149   164       BY SIMILARITY.
FT   DISULFID    166   175       BY SIMILARITY.
FT   DISULFID    181   222       BY SIMILARITY.
FT   DISULFID    208   235       BY SIMILARITY.
FT   DISULFID    240   285       BY SIMILARITY.
FT   DISULFID    271   298       BY SIMILARITY.
FT   DISULFID    303   348       BY SIMILARITY.
FT   DISULFID    334   361       BY SIMILARITY.
FT   DISULFID    366   407       BY SIMILARITY.
FT   DISULFID    393   420       BY SIMILARITY.
FT   CARBOHYD     61    61       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD     65    65       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     79    79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    160   160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201   201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    254   254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    376   376       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400   400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    253   253       C -> Y (IN REF. 2).
FT   CONFLICT    313   313       L -> F (IN REF. 2).
FT   CONFLICT    321   321       T -> N (IN REF. 2).
FT   CONFLICT    327   327       K -> N (IN REF. 2).
FT   CONFLICT    363   363       V -> A (IN REF. 2).
FT   CONFLICT    384   384       V -> M (IN REF. 2).
FT   CONFLICT    461   484       KFVPSSSSECLQPNGSYQMPSDLI ->
FT                               NLELPAAPNAFNPMDPTKCLLF (IN REF. 2).
SQ   SEQUENCE   484 AA;  52567 MW;  AFF74F625C1FD013 CRC64;

  Query Match             16.4%;  Score 123.5;  DB 1;  Length 484;
  Best Local Similarity   23.8%;  Pred. No. 4.6e-05;
  Matches   34;  Conservative   18;  Mismatches   62;  Indels   29;  Gaps    5;

Qy       2 ISCDPPPPVKNARKPYYSLPIVP-----GIVLRYTCSPSYRLIGEKAIFCISENQVHAIW 56
           : |:   |     |   :   |         |   : |   : ||| : : | |      :|
Db     173 LQCEQVVECDALENPVNGVVICPQSLPWNTTCAFECKEGFELIGPEHLQCTSS----GSW 228

Qy      57 DKAPPICESV---------NKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFIMK 107
           |   | |::|         |  :||          | |   | :  :  |||   | ::
Db     229 DGKKPTCKAVTCDTVGHPQNGDVSC----------NHSSIGEPAYKSTCHFTCAEGFQLQ 278

Qy     108 GSKTVWCQANEMWGPTALPVCES 130
           |   : | |   |   | ||::
Db     279 GPAQIECIAQGQWTQQA-PVCKA 300


RESULT 43
APOE_CANFA
ID   APOH_CANFA      STANDARD;       PRT;   345 AA.
AC   P33703;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI).
```

```
GN   APOH
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BEAGLE; TISSUE=Liver;
RX   MEDLINE=94221500; PubMed=7682067;
RA   Sellar G.C., Keane J., Mehdi H., Peeples M.E., Browne N.,
RA   Whitehead A.S.;
RT   "Characterization and acute phase modulation of canine apolipoprotein
RT   H (beta 2 glycoprotein I).";
RL   Biochem. Biophys. Res. Commun. 191:1288-1293(1993).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X72933; CAA51438.1; -.
DR   PIR; JN0465; JN0465.
DR   HSSP; P02749; 1C1Z.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT   SIGNAL       1     19       BY SIMILARITY.
FT   CHAIN       20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      22     80       SUSHI 1.
FT   DOMAIN      83    138       SUSHI 2.
FT   DOMAIN     141    203       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     263    345       SUSHI-LIKE.
FT   DISULFID    23     66       BY SIMILARITY.
FT   DISULFID    51     79       BY SIMILARITY.
FT   DISULFID    84    124       BY SIMILARITY.
FT   DISULFID   110    137       BY SIMILARITY.
FT   DISULFID   142    188       BY SIMILARITY.
FT   DISULFID   174    200       BY SIMILARITY.
FT   DISULFID   205    248       BY SIMILARITY.
FT   DISULFID   244    260       BY SIMILARITY.
FT   DISULFID   264    315       BY SIMILARITY.
FT   DISULFID   300    329       BY SIMILARITY.
FT   DISULFID   307    345       BY SIMILARITY.
FT   CARBOHYD   117    117       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   162    162       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   183    183       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   193    193       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   253    253       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   345 AA;  38403 MW;  E0B262487B874FEA CRC64;

  Query Match             16.4%;  Score 124;  DB 1;  Length 345;
  Best Local Similarity   26.1%;  Pred. No. 3.6e-05;
  Matches   36;  Conservative   18;  Mismatches   68;  Indels   16;  Gaps    5;

Qy       2 ISCPEPPEVKNAR    KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           ::| ||   | |      ||  :  : | | | :  | | :    | |
Db     140 VTCPFPSVPKFAILSVFKPLATNNSLYGNKAVFECLPHYAMPGNDTITCTA--BGNWT 195

Qy      58 KAPPICESVNKTISCSDPTVP-GGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQA 116
             |  |     : |  |  |  |||  :|    : |   : |   :|: | : | |
Db     196 ILPECRE     VKCPFPSRPDNGFVNYPAKPTLYYKDKAMYGCHDTYTLDGPEVVECNK 250

Qy     117 NEMWGPTALPVCESDFPL 134
             |  :| | |::   |
Db     251 FGNW  SAQPSCKASCKL 266


RESULT 44
DAF_CAVPO
ID   DAF_CAVPO      STANDARD;      PRT;   507 AA.
AC   Q60401; Q60402; Q60403; Q60404; Q60405; Q60406; P97254; P97255;
AC   P97256; Q9WTI9;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Complement decay-accelerating factor precursor.
GN   DAF.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=HARTLEY; TISSUE=Spleen;
RX   MEDLINE=95403978; PubMed=7545710;
RA   Nonaka M., Miwa T., Okada N., Nonaka M., Okada H.;
RT   "Multiple isoforms of guinea pig decay-accelerating factor (DAF)
RT   generated by alternative splicing.";
RL   J. Immunol. 155:3037-3048(1995).
CC   -!- FUNCTION: PREVENTS THE FORMATION AND/OR ACCELERATES THE
CC       DISSOCIATION OF C3 CONVERTASE.
CC   -!- SUBCELLULAR LOCATION: CLASS GPI: ATTACHED TO THE MEMBRANE BY A
CC       GPI-ANCHOR.
CC   -!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST SIX FORMS OF DAF
CC       TRANSCRIPTS (DISTRIBUTED IN FOUR CLASSES: GPI, TCL, TCS AND SEC)
CC       GENERATED BY ALTERNATIVE SPLICING. ONE GPI ANCHORED FORM (GDAB-
CC       GPI), TWO FORMS WITH LONGER CYTOPLASMIC REGION (GDAB-TCL AND
CC       GDABC-TCL), TWO FORMS WITH SHORTER CYTOPLASMIC REGION (GDA-TCS AND
CC       GDAB-TCS) AND ONE SECRETED FORM (GDAB-SEC).
CC   -!- TISSUE SPECIFICITY: ALL THE ISOFORMS ARE WIDELY EXPRESSED. GPI AND
CC       TCS ARE THE MAJOR FORMS, WHEREAS SEC IS MINOR AND TCL IS ONLY
CC       PRESENT IN TRACE LEVELS.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC       (RCA) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D55667; BAA09514.1; -.
DR   EMBL; D55656; BAA09514.1; JOINED.
DR   EMBL; D55657; BAA09514.1; JOINED.
DR   EMBL; D55658; BAA09514.1; JOINED.
DR   EMBL; D55659; BAA09514.1; JOINED.
DR   EMBL; D55660; BAA09514.1; JOINED.
DR   EMBL; D55661; BAA09514.1; JOINED.
DR   EMBL; D55662; BAA09514.1; JOINED.
DR   EMBL; D55663; BAA09514.1; JOINED.
DR   EMBL; D55664; BAA09514.1; JOINED.
DR   EMBL; D55667; BAA09515.1; -.
DR   EMBL; D55656; BAA09515.1; JOINED.
DR   EMBL; D55657; BAA09515.1; JOINED.
DR   EMBL; D55658; BAA09515.1; JOINED.
DR   EMBL; D55659; BAA09515.1; JOINED.
DR   EMBL; D55660; BAA09515.1; JOINED.
DR   EMBL; D55661; BAA09515.1; JOINED.
DR   EMBL; D55662; BAA09515.1; JOINED.
DR   EMBL; D55663; BAA09515.1; JOINED.
DR   EMBL; D55664; BAA09515.1; JOINED.
DR   EMBL; D55665; BAA09515.1; JOINED.
```

```
DR   EMBL; D55667; BAA09516.1; -.
DR   EMBL; D55656; BAA09516.1; JOINED.
DR   EMBL; D55657; BAA09516.1; JOINED.
DR   EMBL; D55658; BAA09516.1; JOINED.
DR   EMBL; D55659; BAA09516.1; JOINED.
DR   EMBL; D55660; BAA09516.1; JOINED.
DR   EMBL; D55661; BAA09516.1; JOINED.
DR   EMBL; D55662; BAA09516.1; JOINED.
DR   EMBL; D55663; BAA09516.1; JOINED.
DR   EMBL; D55664; BAA09516.1; JOINED.
DR   EMBL; D55665; BAA09516.1; JOINED.
DR   EMBL; D55667; BAA09517.1; -.
DR   EMBL; D55656; BAA09517.1; JOINED.
DR   EMBL; D55657; BAA09517.1; JOINED.
DR   EMBL; D55658; BAA09517.1; JOINED.
DR   EMBL; D55659; BAA09517.1; JOINED.
DR   EMBL; D55660; BAA09517.1; JOINED.
DR   EMBL; D55661; BAA09517.1; JOINED.
DR   EMBL; D55662; BAA09517.1; JOINED.
DR   EMBL; D55663; BAA09517.1; JOINED.
DR   EMBL; D55664; BAA09517.1; JOINED.
DR   EMBL; D55665; BAA09517.1; JOINED.
DR   EMBL; D55666; BAA09517.1; JOINED.
DR   EMBL; D55667; BAA09518.1; -.
DR   EMBL; D55656; BAA09518.1; JOINED.
DR   EMBL; D55657; BAA09518.1; JOINED.
DR   EMBL; D55658; BAA09518.1; JOINED.
DR   EMBL; D55659; BAA09518.1; JOINED.
DR   EMBL; D55660; BAA09518.1; JOINED.
DR   EMBL; D55661; BAA09518.1; JOINED.
DR   EMBL; D55662; BAA09518.1; JOINED.
DR   EMBL; D55663; BAA09518.1; JOINED.
DR   EMBL; D55664; BAA09518.1; JOINED.
DR   EMBL; D55665; BAA09518.1; JOINED.
DR   EMBL; D55666; BAA09518.1; JOINED.
DR   EMBL; D55667; BAA09519.1; -.
DR   EMBL; D55656; BAA09519.1; JOINED.
DR   EMBL; D55657; BAA09519.1; JOINED.
DR   EMBL; D55658; BAA09519.1; JOINED.
DR   EMBL; D55659; BAA09519.1; JOINED.
DR   EMBL; D55660; BAA09519.1; JOINED.
DR   EMBL; D55661; BAA09519.1; JOINED.
DR   EMBL; D55662; BAA09519.1; JOINED.
DR   EMBL; D55663; BAA09519.1; JOINED.
DR   EMBL; D55664; BAA09519.1; JOINED.
DR   EMBL; D55666; BAA09519.1; JOINED.
DR   EMBL; D49416; BAA08396.1; -.
DR   EMBL; D49417; BAA08397.1; -.
DR   EMBL; D49418; BAA08398.1; -.
DR   EMBL; D49419; BAA08399.1; -.
DR   EMBL; D49420; BAA08400.1; -.
DR   EMBL; D49421; BAA08401.1; -.
DR   HSSP; P08603; 1HCC.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
KW   Complement pathway; Glycoprotein; Membrane; Repeat; GPI anchor;
KW   Alternative splicing; Signal; Sushi.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    451       COMPLEMENT DECAY-ACCELERATING FACTOR.
FT   PROPEP      452    507       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       32     93       SUSHI 1.
FT   DOMAIN       94    158       SUSHI 2.
FT   DOMAIN      159    220       SUSHI 3.
FT   DOMAIN      221    283       SUSHI 4.
FT   DOMAIN      284    458       SER/THR-RICH.
FT   LIPID       451    451       GPI-ANCHOR (BY SIMILARITY).
FT   CARBOHYD     57     57       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC      ) (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    493    493       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    302    344       MISSING (IN ISOFORM GDA-TCS).
FT   VARSPLIC    335    366       MISSING (IN ISOFORM GDAB-GPI, ISOFORM
FT                                GDAB-TCL, ISOFORM GDA-TCS, ISOFORM GDAB
FT                                TCS AND ISOFORM GDAB-SEC).
FT   VARSPLIC    399    424       MISSING (IN ISOFORM GDAB-GPI AND GDAB
FT                                SEC).
FT   VARSPLIC    425    458       MISSING (IN ISOFORM GDAB-GPI, ISOFORM
FT                                GDAB-TCL, ISOFORM GDABC-TCL AND ISOFORM
FT                                GDAB-SEC).
FT   VARSPLIC    459    507       THVYKVLSEACGASNHWLADIAKEDLRRDFSNAQNISSLLQ
FT                                VLGAAQTQ -> GHMCIKLTVLLVVVLVTIG (IN
FT                                ISOFORM GDAB-GPI).
FT   VARSPLIC    459    507       THVYKVLSEACGASNHWLADIAKEDLRRDFSNAQNISSLLQ
FT                                VLGAAQTQ -> DTCV (IN ISOFORM GDAB-SEC).
FT   VARSPLIC    459    507       THVYKVLSEACGASNHWLADIAKEDLRRDFSNAQNISSLLQ
FT                                VLGAAQTQ -> ANMRIMRTKNPMLYVIT (IN ISOFORM
FT                                GDA-TCS AND ISOFORM GDAB-TCS).
SQ   SEQUENCE   507 AA;  55263 MW;  D25BBB774942E210 CRC64;

  Query Match             16.1%;  Score 121;  DB 1;  Length 507;
  Best Local Similarity   26.1%;  Pred. No. 8.6e-05;
  Matches   35;  Conservative   21;  Mismatches   66;  Indels   12;  Gaps    6;

Qy       4 CDPPPEVKNARKPYYSLPIVPGTVLRYTCSESYRLIGEKAIFCISENQVHA WDKAPPIC 63
           |  | |: |       :  :: |: : ::| | ||| || : ||: :      |  : | |
Db     162 CPNPGELVNG-NINVTTDILLGSQIFFSCDEGYRLTGEASAFCMIKGNA VGWSSSLPTC 219

Qy      64 ESVNKTISCSDP--IVPGGFMNKGSKAPFPHGLSVTFTCKANFTMKGSKTVWCQANEMWG 121
             | | :|  |  |  :|:     :||   ||: |   | : |   :  |   :  |
Db     220 IK----LICPEPPQIENGRIVNEEDTYEYRH--VVTYFCNKGFVLTGKSHISCIVRLDVG 273

Qy     122 PTA--LPVCESDFP 133
             :  | |  |
Db     274 EWSDPPPTCRVKSP 287


RESULT 45
APOH_HUMAN
ID   APOH_HUMAN      STANDARD;       PRT;   345 AA.
AC   P02749;
DT   21-JUL 1986 (Rel. 01, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE   (Beta(2)GPI) (Activated protein C-binding protein) (APC inhibitor).
GN   APOH OR B2GI
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver
RX   MEDLINE=91315408; PubMed=1650181;
RA   Steinkasserer A., Estaller C., Weiss E., Sim R.B., Day A.J.;
RT   "Complete nucleotide and deduced amino acid sequence of human beta 2-
RT   glycoprotein I.";
RL   Biochem. J. 277:387-391(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver
RX   MEDLINE=92008618; PubMed=1655523;
RA   Kristensen T., Schousboe I., Boel E., Mulvihill E.M., Hansen P.R.,
RA   Moeller K.B., Moeller N.P.H., Sottrup-Jensen L.;
RT   "Molecular cloning and mammalian expression of human beta
RT   2-glycoprotein I cDNA.";
RL   FEBS Lett. 289:183-186(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92081151; PubMed=1748314;
RA   Mehdi H., Nunn M., Steel D.M., Whitehead A.S., Perez M., Walker I.,
RA   Peeples M.E.;
RT   "Nucleotide sequence and expression of the human gene encoding
RT   apolipoprotein H (beta 2-glycoprotein I).";
```

```
RL   Gene 108:293-298(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92273779; PubMed=1339416;
RA   Day J.R., O'Hara P.J., Grant F.J., Lotton Day C.E., Berkaw M.N.,
RA   Wermer P., Armand P.;
RT   "Molecular cloning and sequence analysis of the cDNA encoding human
RT   apolipoprotein H (beta 2 glycoprotein I).";
RL   Int. J. Clin. Lab. Res. 21:256-264(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92185065; PubMed=1737418;
RA   Matsuura E., Igarashi M., Igarashi Y., Nagae R., Ichikawa K.,
RA   Yasuda T., Koike T.;
RT   "Molecular definition of human beta 2-glycoprotein I (beta 2-GPI) by
RT   cDNA cloning and inter-species differences of beta 2-GPI in
RT   alternation of anticardiolipin binding.";
RL   Int. Immunol. 4:1217-1221(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99115472; PubMed=9914524;
RA   Ohkubo H., Pasmussen J.E., Sanghera D.K., Kamboh M.I., Kristensen T.;
RT   "Structure of the human beta2-glycoprotein I (apolipoprotein H)
RT   gene.";
RL   Eur. J. Biochem. 259:439-440(1999).
RN   [6]
RP   SEQUENCE OF 20-345, CARBOHYDRATE LINKAGE SITES, AND DISULFIDE BONDS.
RX   MEDLINE=84222015; PubMed=6587378;
RA   Lozier J., Takahashi N., Putnam F.W.;
RT   "Complete amino acid sequence of human plasma beta 2-glycoprotein I.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:3640-3644(1984).
RN   [7]
RP   SEQUENCE OF 20-48.
RC   TISSUE=Follicular fluid;
RX   MEDLINE=21148189; PubMed=11250549;
RA   Aleporou-Marinou V., Pappa H., Yaloutis P., Patargias T.;
RT   "Purification of apolipoprotein H (beta 2 glycoprotein I) like protein
RT   from human follicular fluid.";
RL   Comp. Biochem. Physiol. 128B:537-542(2001).
RN   [8]
RP   DISULFIDE BONDS IN C-TERMINAL DOMAIN.
RX   MEDLINE=90050249; PubMed=1426288;
RA   Steinkasserer A., Barlow P.N., Willis A.C., Kertesz Z.,
RA   Campbell I.D., Sim R.B., Norman D.G.;
RT   "Activity, disulphide mapping and structural modelling of the fifth
RT   domain of human beta 2-glycoprotein I.";
RL   FEBS Lett. 314:198-197(1992).
RN   [9]
RP   STRUCTURE OF CARBOHYDRATES.
RX   MEDLINE=97299942; PubMed=9150091;
RA   Gambino R., Ruin G., Pagano G., Cassader M.;
RT   "Qualitative analysis of the carbohydrate composition of
RT   apolipoprotein H.";
RL   J. Protein Chem. 16:205-212(1997).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RC   TISSUE=Plasma;
RX   MEDLINE=99437994; PubMed=10508150;
RA   Bouma B., de Groot P.G., van Den Elsen J.M.H., Ravelli R.B.G.,
RA   Schouten A., Simmelink M.J.A., Derksen R.H.W.M., Kroon J., Gros P.;
RT   "Adhesion mechanism of human beta(2)-glycoprotein I to phospholipids
RT   based on its crystal structure.";
RL   EMBO J. 18:5166-5174(1999).
RN   [12]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX   MEDLINE=20031644; PubMed=10562535;
RA   Schwarzenbacher R., Zeth K., Diederichs K., Gries A., Kostner G.M.,
RA   Laggner P., Prassl R.;
RT   "Crystal structure of human beta2-glycoprotein I: implications for
RT   phospholipid binding and the antiphospholipid syndrome.";
RL   EMBO J. 18:6228-6239(1999).
RN   [13]
RP   VARIANT TYR 256.
RX   MEDLINE=95273813; PubMed=8099061;
RA   Steinkasserer A., Doerner C., Wuerzner R., Sim R.B.;
RT   "Human beta 2-glycoprotein I: molecular analysis of DNA and amino
RT   acid polymorphism.";
RL   Hum. Genet. 91:401-402(1993).
RN   [14]
RP   VARIANT ASN 107.
RX   MEDLINE=97369481; PubMed=9225969;
RA   Sanghera D.K., Kristensen T., Hamman R.F., Kamboh M.I.;
RT   "Molecular basis of the apolipoprotein H (beta 2-glycoprotein I)
RT   protein polymorphism.";
RL   Hum. Genet. 100:57-62(1997).
RN   [15]
RP   VARIANTS GLY 325 AND SER 335.
RX   MEDLINE 97217791; PubMed 9064752;
RA   Sanghera D.K., Wagenknecht D.R., McIntyre J.A., Kamboh M.I.;
RT   "Identification of structural mutations in the fifth domain of
RT   apolipoprotein H (beta-2 glycoprotein I) which affect phospholipid
RT   binding.";
RL   Hum. Mol. Genet. 6:411-416(1997).
CC   -!- FUNCTION: BINDS TO VARIOUS KINDS OF NEGATIVELY CHARGED SUBSTANCES
CC       SUCH AS HEPARIN, PHOSPHOLIPIDS, AND DEXTRAN SULFATE. MAY PREVENT
CC       ACTIVATION OF THE INTRINSIC BLOOD COAGULATION CASCADE BY BINDING
CC       TO PHOSPHOLIPIDS ON THE SURFACE OF DAMAGED CELLS.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
CC   ---------- ------------------------- -- -- -- -- -- -- -- -- -- -- -- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------- ------------------------- -- -- -- -- -- -- -- -- -- -- -- --
DR   EMBL; X58100; CAA41113.1; -.
DR   EMBL; X53595; CAA37664.1; -.
DR   EMBL; X57847; CAA40977.1; -.
DR   EMBL; M62839; AAA51766.1; -.
DR   EMBL; S80305; AAB21330.1; -.
DR   EMBL; Y11493; CAA72279.1; -.
DR   EMBL; Y11494; CAA72279.1; JOINED.
DR   EMBL; Y11495; CAA72279.1; JOINED.
DR   EMBL; X53595; CAA72279.1; JOINED.
DR   EMBL; Y11496; CAA72279.1; JOINED.
DR   EMBL; Y11497; CAA72279.1; JOINED.
DR   EMBL; Y11498; CAA72279.1; JOINED.
DR   EMBL; Y17754; CAA76845.1; -.
DR   PIR; S17178; NHHU.
DR   PIR; B43286; B43286.
DR   PDB; 1QUB; 08-OCT-99.
DR   PDB; 1C1Z; 19-NOV-99.
DR   MIM; 138700; -.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal;
KW   Polymorphism; 3D-structure.
FT   SIGNAL       1     19
FT   CHAIN       20    345       BETA-2-GLYCOPROTEIN I.
FT   DOMAIN      22     80       SUSHI 1.
FT   DOMAIN      83    138       SUSHI 2.
FT   DOMAIN     141    201       SUSHI 3.
FT   DOMAIN     204    261       SUSHI 4.
FT   DOMAIN     263    345       SUSHI-LIKE.
FT   DISULFID    23     66
FT   DISULFID    51     79
FT   DISULFID    84    124
FT   DISULFID   110    137
FT   DISULFID   142    188
FT   DISULFID   174    200
FT   DISULFID   205    248
FT   DISULFID   234    260
```

```
FT   DISULFID     264     315
FT   DISULFID     300     325
FT   DISULFID     307     345
FT   CARBOHYD     149     149         O-LINKED.
FT   CARBOHYD     162     162         N-LINKED (GLCNAC. . .).
FT   CARBOHYD     183     183         N-LINKED (GLCNAC. . .).
FT   CARBOHYD     193     193         N-LINKED (GLCNAC. . .).
FT   CARBOHYD     253     253         N-LINKED (GLCNAC. . .).
FT   VARIANT      107     107         S -> N (IN APOH*1).
FT                                    /FTId=VAR_008169.
FT   VARIANT      266     266         V -> L (IN 23% OF THE POPULATION).
FT                                    /FTId=VAR_000673.
FT   VARIANT      325     325         C -> G (LOSS OF PHOSPHATIDYLSERINE-
FT                                    BINDING).
FT                                    /FTId=VAR_008170.
FT   VARIANT      335     335         W -> S (IN APOH*3W; LOSS OF
FT                                    PHOSPHATIDYLSERINE-BINDING).
```

```
  Query Match             16.0%;  Score 120;  DB 1;  Length 345;
  Best Local Similarity   25.4%;  Pred. No. 7.1e-05;
  Matches   44;  Conservative   19;  Mismatches   65;  Indels   16;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLR---YTCSPSYRLIGEKAIFCISENQVHATWD 57
           | | ||     |    |       :: |    : | | : : |   | | :    | | |
Db     140 IICPPPSIPTFATLRVYKPSAGNNSLYRDIAVFECLPQHAMFGNDTITCTT----HGNWT 195

Qy      58 KAPPICESVNKTISCSDPIVP-GGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQA 116
           | |   |     : |  |  |  ||:|  :|    : |  || |   ::: | : : |
Db     196 KLPECRE----VKCPFPSRPDNGFVNYPAKPTLYYKDKATFGCHDGYSLDGPEEIECTK 250

Qy     117 NEMWGPTALPVCES 130
             |  :|:| |::
Db     251 LGNW-SAMPSCKA 262


Search completed: November  6, 2002, 16:05:32
Job time : 13.5816 secs
```

OM protein - protein search, using sw model

Run on: November 6, 2002, 16:05:00 ; Search time 21.5799 Seconds
(without alignments)
1082.224 Million cell updates/sec

Title: US-09-834-309-6
Perfect score: 752
Sequence: 1 EISCDPPPEVKNARKPYYSI ... ANEMWGPTALPVCESDFPLE 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	752	100.0	1032	11	Q9dc83	Q9dc83 mus musculu
2	749	99.6	679	11	Q99254	Q99254 mus musculu
3	499	66.4	1045	6	O46545	O46545 ovis aries
4	466.5	62.0	1032	4	Q13866	Q13866 homo sapien
5	466.5	62.0	1087	4	Q14212	Q14212 homo sapien
6	241.5	32.8	1911	6	Q29528	Q29528 papio hamad
7	224	29.8	533	11	Q08569	Q08569 cavia porce
8	222	29.5	469	11	Q91X48	Q91x48 mus musculu
9	221.5	29.5	2014	6	Q29530	Q29530 pan troglod
10	221.5	29.5	2039	4	Q16745	Q16745 homo sapien
11	221.5	29.5	2489	4	Q16744	Q16744 homo sapien
12	220.5	29.3	132	4	Q9HB99	Q9hb99 homo sapien
13	220.5	29.3	132	4	Q9HB98	Q9hb98 homo sapien
14	219.5	29.2	132	4	Q9HC55	Q9hc55 homo sapien
15	215.5	28.7	142	4	Q9HC54	Q9hc54 homo sapien
16	215	28.6	579	11	Q60736	Q60736 mus musculu
17	214	28.5	555	11	Q99JA1	Q99ja1 cavia porce
18	199.5	26.5	363	6	Q02839	Q02839 sus scrofa
19	198.5	26.4	300	11	Q9R0R5	Q9r0r9 mus musculu
20	198.5	26.4	365	11	Q88174	Q88174 mus musculu
21	197.5	26.3	263	12	Q89859	Q89859 variola vir
22	197.5	26.3	263	12	Q07035	Q67035 variola vir
23	195.5	26.0	263	12	Q89076	Q89076 variola vir
24	195.5	26.0	263	12	Q89061	Q89061 variola vir
25	193.5	25.7	259	12	P87616	P87616 cowpox viru
26	188.5	25.1	349	4	Q15429	Q15429 homo sapien
27	185.5	24.7	645	12	Q9WRU2	Q9wru2 macaca mula
28	184.5	24.5	285	6	O19121	O19121 papio hamad
29	184.5	24.5	1001	4	Q9Y2E1	Q9y2e1 homo sapien
30	183.5	24.4	661	6	Q29531	Q29531 pan troglod
31	182.5	24.3	315	6	Q28770	Q28770 papio cynoc
32	181.5	24.1	285	6	O19126	O19126 macaca fasc
33	181.5	24.1	285	6	O19127	O19127 macaca mula
34	181.5	24.1	369	6	P79138	P79138 cercopithec
35	178.5	23.7	559	4	Q9UQV2	Q9uqv2 homo sapien
36	177	23.5	451	13	Q9DEG0	Q9deg0 gallus gall
37	175.5	23.3	222	6	O19120	O19120 aotus trivi
38	175.5	23.3	222	6	O19125	O19125 saimiri sci
39	175.5	23.3	336	6	O62834	O62834 saguinus oe
40	174	23.1	355	11	Q920M4	Q920m4 rattus norv
41	173	23.0	3567	11	Q9ES77	Q9es77 mus musculu
42	172.5	22.9	417	11	O35520	O35520 rattus norv
43	172.5	22.9	497	11	Q63612	Q63612 rattus norv
44	172.5	22.9	559	11	Q63135	Q63135 rattus norv
45	172	22.9	1236	11	Q91YB6	Q91yb6 rattus norv
46	171.5	22.8	550	12	P88903	P88903 kaposi's sa
47	170.5	22.7	222	6	O19122	O19122 callimico g
48	170	22.6	395	12	Q972M6	Q972m6 macaca mula
49	168.5	22.4	222	6	O19124	O19124 saguinus my
50	168.5	22.4	314	6	O62835	O62835 saguinus oe
51	168.5	22.4	377	6	O62838	O62838 saguinus oe
52	168.5	22.4	378	6	O62837	O62837 saguinus oe
53	168	22.3	808	11	Q61408	Q61408 mus musculu
54	167.5	22.3	222	6	O19128	O19128 pithecia pi
55	167.5	22.3	225	12	Q91MM3	Q91mm3 lumpy skin
56	167.5	22.3	550	12	Q40912	Q40912 kaposi's sa
57	167	22.2	483	11	Q64735	Q64735 mus musculu
58	166.5	22.1	522	6	Q28769	Q28769 papio cynoc
59	164.5	21.9	267	11	P97939	P97939 cavia porce
60	164.5	21.9	267	12	Q9DHG9	Q9dhg9 yaba like d
61	164.5	21.9	337	11	P97263	P97263 cavia porce
62	164.5	21.9	339	11	P97261	P97261 cavia porce
63	164.5	21.9	379	11	P70105	P70105 cavia porce
64	164.5	21.9	1653	5	Q9VIU9	Q9viu9 drosophila
65	163.5	21.7	522	6	Q28797	Q28797 pan troglod
66	162.5	21.6	300	12	Q9Q8F9	Q9q8f9 myxoma viru
67	162	21.5	711	4	Q9NUI5	Q9nui5 homo sapien
68	162	21.5	716	4	O95917	O95917 homo sapien
69	161	21.4	777	4	Q9NUI3	Q9nui3 homo sapien
70	161	21.4	787	4	Q9NUI4	Q9nui4 homo sapien
71	161	21.4	792	4	Q9Y3J6	Q9y3j6 homo sapien
72	161	21.4	1024	4	Q9BYH1	Q9byh1 homo sapien
73	157.5	20.9	669	6	Q28085	Q28085 bos taurus
74	157	20.9	268	12	Q9QAX6	Q9qax6 yaba monkey
75	156.5	20.8	251	4	Q9BS25	Q9bs25 homo sapien
76	155.5	20.7	222	6	O19123	O19123 callithrix
77	155	20.6	3564	11	Q923L3	Q923l3 mus musculu
78	153.5	20.4	754	13	Q98C16	Q98c16 cyprinus ca
79	153.5	20.4	754	4	Q96HX6	Q96hx6 homo sapien
80	153	20.3	569	4	Q9BXR6	Q9bxr6 homo sapien
81	151	20.1	923	11	Q91X64	Q91x64 mus musculu
82	150	19.9	330	11	Q62479	Q62479 mus musculu
83	150	19.9	359	13	O73756	O73756 brachydanio
84	149.5	19.9	194	11	Q9QYQ8	Q9qyq8 rattus norv
85	149	19.8	390	11	Q921P0	Q921p0 mus musculu
86	149	19.8	657	4	Q14006	Q14006 homo sapien
87	149	19.8	809	4	Q9HW82	Q9hw82 homo sapien
88	149	19.8	840	4	Q9UJ45	Q9uj45 homo sapien
89	149	19.8	853	4	Q9UJ47	Q9uj47 homo sapien

90	149	14.8	1172	4	Q9N0B7	Q9n0b7 homo sapien
91	148.5	14.7	787	13	Q90422	Q90422 brachydanio
92	148	14.7	699	11	Q91WX0	Q91wx0 rattus norv
93	146.5	14.5	475	4	Q96NI5	Q96ni5 homo sapien
94	145.5	14.4	216	12	Q98VI5	Q98vi5 monkeypox v
95	145.5	14.4	3389	4	Q96QU9	Q96qu9 homo sapien
96	145.5	14.4	3508	4	Q96RM4	Q96rm4 homo sapien
97	145	14.4	472	11	Q9QYJ9	Q9qyj9 rattus norv
98	145	14.4	482	11	Q9QUT3	Q9qut3 rattus norv
99	145	14.4	499	11	Q9QUN6	Q9qun6 rattus norv
100	145	14.4	899	11	Q9Z0M0	Q9z0m0 rattus norv

ALIGNMENTS

RESULT 1
Q9D6Z3
ID Q9D6Z3 PRELIMINARY; PRT; 1042 AA.
AC Q9D6Z3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE COMPLEMENT RECEPTOR 2.
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SPLEEN;
RX MEDLINE=21085600; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kouchi P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bottelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Buti C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK003072; BAB22548.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 14.
DR SMART; SM00032; CCP; 14.
SQ SEQUENCE 1042 AA; 114335 MW; 6F96C4D95E431E89 CRC64;

Query Match 100.0%; Score 792; DB 11; Length 1042;
Best Local Similarity 100.0%; Prod. No. 1.7e-76;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
Db 20 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 79

Qy 61 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 120
Db 80 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 139

Qy 121 GPTALPVCESDFPLE 135
Db 140 GPTALPVCESDFPLE 154

RESULT 2
Q99254
ID Q99254 PRELIMINARY; PRT; 679 AA.
AC Q99254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90229754; PubMed=2139460;
RA Kurtz C.B., O'Toole E., Christenson S.M., Wois J.B.;
RT "The murine complement receptor gene family. IV. Alternative splicing
RT of Cr2 gene transcripts predicts two distinct gene products that share
RT homologous domains with both human CR2 and CR1.";
RL J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE=95105691; PubMed=7528766;
RA Kim Y.U., Kinoshita T., Molina H., Bourcade D., Seya T., Wagner L.M.,
RA Holers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
RT mechanisms of both human decay-accelerating factor and membrane
RT cofactor protein.";
RL J. Exp. Med. 181:151-159(1995).
DR EMBL; U17128; AAA78271.1; -.
DR EMBL; U17123; AAA78271.1; JOINED.
DR EMBL; U17124; AAA78271.1; JOINED.
DR EMBL; U17125; AAA78271.1; JOINED.
DR EMBL; U17126; AAA78271.1; JOINED.
DR EMBL; U17127; AAA78271.1; JOINED.
DR EMBL; M36470; AAA37449.1; -.
DR HSSP; P10998; 1VVD.
DR MGD; MGI:88489; Cr2.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 10.
DR PRINTS; PR00434; SELECTIN.
DR SMART; SM00032; CCP; 10.
FT NON_TER 679 679
SQ SEQUENCE 679 AA; 74916 MW; 52FC00E1CED20CDC CRC64;

Query Match 99.6%; Score 749; DB 11; Length 679;
Best Local Similarity 99.3%; Prod. No. 2.3e-76;
Matches 134; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 60
Db 396 QISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAP 455

Qy 61 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 120
Db 456 PICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMW 515

Qy 121 GPTALPVCESDFPLE 135
Db 516 GPTALPVCESDFPLE 530

RESULT 3
O46545
ID O46545 PRELIMINARY; PRT; 1045 AA.

```
AC   O46545;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 2.
GN   CR2.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE ALPINE;
RX   MEDLINE=98309471; PubMed=9647256;
RA   Hein W.R., Dudler L., Marston W.L., Landsverk T., Young A.J.,
RA   Avila D.;
RT   "Ubiquitination and dimerization of complement receptor type 2 on
RT   sheep B cells.";
RL   J. Immunol. 161:458-466(1998).
DR   EMBL; AF038131; AAB92375.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 14.
DR   SMART; SM00032; CCP; 14.
KW   Receptor.
SQ   SEQUENCE   1045 AA;  115562 MW;  FF58E1A2892C0D59 CRC64;

  Query Match             66.4%;  Score 499;  DB 6;  Length 1045;
  Best Local Similarity   61.2%;  Pred. No. 8e-48;
  Matches   82;  Conservative   22;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           | ||||| :|| |  |:| |:    |: |:|: ::|||||: :|||||: |:  ||||||
Db      40 IFCDPPPSIKNGRSGYHSGPVALNAVVTYSCASAFRLIGERRLFCISKDNVNGIWDKAPP 89

Qy      62 ICFSVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           ||| |:   ||:||:|||:  | |: ||||||:|||:| |||||||:||||||||  ||
Db      90 ICEYYNRNSVCSEPIIPGGYRIKISKIPFFRHGDAVIFSCNANFTMKGNKIVWCQANRKWG 149

Qy     122 PTALPVCESDFPLE 135
           || ||:|||| |||
Db     150 PTPLPICESDTPLE 163


RESULT 4
Q13866
ID   Q13866       PRELIMINARY;      PRT;  1032 AA.
AC   Q13866;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CR2 RECEPTOR PRECURSOR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-LYMPHOCYTES OF TONSILS;
RX   MEDLINE=88171282; PubMed=2832596;
RA   Weis J.J., Toothaker L.E., Smith J.A., Weis J.H., Fearon D.T.;
RT   "Structure of the human B lymphocyte receptor for C3d and the Epstein
RT   Barr virus and relatedness to other members of the family C3/C4
RT   binding proteins.";
RL   J. Exp. Med. 167:1047-1066(1988).
DR   EMBL; Y00649; CAA68674.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 15.
DR   SMART; SM00032; CCP; 14.
DR   PROSITE; PS00134; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW   Signal; Receptor.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1032       POTENTIAL.
SQ   SEQUENCE   1032 AA;  112844 MW;  14F7D92CA6A18888 CRC64;

  Query Match             62.0%;  Score 466.5;  DB 4;  Length 1032;
  Best Local Similarity   61.2%;  Pred. No. 3.8e-44;
  Matches   82;  Conservative   18;  Mismatches   33;  Indels    1;  Gaps    1;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           ||| || :  | |  ||| ||  |||:||: | ::|||||:: ||:::|  |||| |
Db      21 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGIFRLIGEKSLLCITKDKVDGIWDKPAP 80

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           ||  ||  || :||||||:  :||  |:| || ||| || ||:| |:|:|||:||| |||
Db      81 KCEYFNKYSSCPEPIVPGGYKIRGS-TPYFHGDSVTFACKTNFSMNGNKSVWCQANNMWG 139

Qy     122 PTALPVCESDFPLE 135
           || || | : ||||
Db     140 PTRLPTCVSVFPLE 153


RESULT 5
Q14212
ID   Q14212       PRELIMINARY;      PRT;  1087 AA.
AC   Q14212;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CR2 PROTEIN PRECURSOR.
GN   CR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88097454; PubMed=2827171;
RA   Moore M., Cooper N., Tack B., Nemerow G.;
RT   "Molecular cloning of the cDNA encoding the Epstein-Barr virus C3d
RT   receptor (complement receptor type 2) of human b lymphocytes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198(1987).
DR   EMBL; J03565; AAA35784.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 16.
DR   SMART; SM00032; CCP; 14.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
KW   Signal.
FT   SIGNAL        1     20       POTENTIAL
FT   CHAIN        21   1087       POTENTIAL
SQ   SEQUENCE   1087 AA;  118933 MW;  CB2B6F350295178B9 CRC64;

  Query Match             62.0%;  Score 466.5;  DB 4;  Length 1087;
  Best Local Similarity   61.2%;  Pred. No. 4e-44;
  Matches   82;  Conservative   18;  Mismatches   33;  Indels    1;  Gaps    1;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           ||| || :  | |  ||| ||  |||:||: | ::|||||:: ||:::|  |||| |
Db      21 ISCGSPPPILNGRISYYSTPIAVGTVIRYSCSGIFRLIGEKSLLCITKDKVDGIWDKPAP 80

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
           ||  ||  || :||||||:  :||  |:| || ||| || ||:| |:|:|||:||| |||
Db      81 KCEYFNKYSSCPEPIVPGGYKIRGS-TPYFHGDSVTFACKTNFSMNGNKSVWCQANNMWG 139

Qy     122 PTALPVCESDFPLE 135
           || || | : ||||
Db     140 PTRLPTCVSVFPLE 153


RESULT 6
Q29528
```

```
ID   Q29528       PRELIMINARY;      PRT;   1911 AA.
AC   Q29528;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT COMPONENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Clemenza I., Subramanian B.V., Nickells M.W., Bourcade D.E.,
RA   Atkinson J.P.;
RT   "Primary sequence of the baboon 200 kDa C4b/C4b receptor (CR1).";
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L39791; AAA62170.1; -.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; Sod_Cu_Zn.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 29.
DR   SMART; SM00032; CCP; 29.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN 4.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN 1.
KW   Receptor.
FT   NON_TER     1911     1911
SQ   SEQUENCE   1911 AA;  210174 MW;  585A4DD0EAFA521D CRC64;

  Query Match             60.8%;  Score 231.5;  DB 6;  Length 1911;
  Best Local Similarity   36.7%;  Pred. No. 3.3e-17;
  Matches   51;  Conservative   37;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA       RKPYYSLPIVPGTVLKYTCSPs -  - YRLIGEKAIFCISE 49
           |||.|||.: :|          ::|     |||: | |          : |:|||:|:| |:
Db    1404 ISCKPPPTISNGDFYSNNRTSFHS  -- GTVVTYQCHTGPDGEQLFFLVGERSIYCTSK 1458

Qy      50 NQVHAIWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPERHGDSVIFICKANFIMKGS 109
           :    |  || : | |    |: | |  |    |::: |  : | | ::  | | ||
Db    1459 DDQVGAWSSPPPRCISTNK  CTAPEVKNGIPVPGNRSFFSINFIVPFPVQPGFVMVGS 1518

Qy     110 KTVWCQANEMWGPTALPVC 128
           || || |  ||| || |
Db    1519 HIVQCQINNRWGP KLPHC 1538


RESULT 7
Q08569
ID   Q08569       PRELIMINARY;      PRT;   533 AA.
AC   Q08569;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ACROSOMAL MATRIX COMPONENT AM67 PRECURSOR.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HARTLEY; TISSUE=TESTIS;
RX   MEDLINE=97284752; PubMed=9139729;
RA   Foster J.A., Friday B.B., Maulit M.T., Blobel C., Wintroy V.P.,
RA   Osmond G.E., Kim K.S., Gerton G.L.;
RT   "AM67, a secretory component of the guinea pig sperm acrosomal matrix,
RT   is related to mouse sperm protein sp56 and the complement component 4
RT   binding proteins.";
RL   J. Biol. Chem. 272:12714-12722(1997).
DR   EMBL; U75654; AAC14888.1; -.
DR   HSSP; P10998; 1VVD
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
KW   Signal.
FT   SIGNAL      1     28       POTENTIAL.
FT   CHAIN      29    533       ACROSOMAL MATRIX COMPONENT AM67.
SQ   SEQUENCE   533 AA;  59772 MW;  EDBDDED487A45389 CRC64;

  Query Match             29.8%;  Score 224;  DB 11;  Length 533;
  Best Local Similarity   30.1%;  Pred. No. 5.4e-17;
  Matches   40;  Conservative   27;  Mismatches   52;  Indels   14;  Gaps    4;

Qy       2 ISCDPPPEVKNAR----KPYYSLPIVPGTVLKYTCSPSYRLIGEKAIFCISENQVHATW 56
           : |: |:: | :     :  |:     |::: | : :: | ::  :: :  |:    |
Db     151 VKCDSPPDISNGRHSGIDEDEYIY    GSLVIYVCDEGYSLLGNASISCLVANETVGVW 206

Qy      57 DKAPPICESVNKTISCSDPIVPGGFMNKGSKAPERHGDSVIFICKANFIMKGSKTVWCQA 116
           || || |     |  | :| |  |    :  : |::  :|  : :|||  : |:|
Db     207 SSNFPICEEV     ICEQPEIFHGIMEFQGREYYYKPEVRYRCKDYYILGGEIEYSCIDGYWA 262

Qy     117 NEMWGPTALPVCE 129
           |  | | ::| ||
Db     263 NSKWYPSIPICE 274


RESULT 8
Q91X48
ID   Q91X48       PRELIMINARY;      PRT;   469 AA.
AC   Q91X48;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT COMPONENT 4 BINDING PROTEIN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC012257; AAH12257.1; -.
SQ   SEQUENCE   469 AA;  51523 MW;  1426683143972197 CRC64;

  Query Match             29.5%;  Score 222;  DB 11;  Length 469;
  Best Local Similarity   34.3%;  Pred. No. 7.9e-17;
  Matches   47;  Conservative   19;  Mismatches   57;  Indels   14;  Gaps    5;

Qy       2 ISCDPPPEVKNAR----KPYYSLPIVPGTVLPYTCSPSYRLIGEKAIFCISENQVHATWD 57
           : | |||:: | :     : :|  |    |  : ||| | :|:|    | :    :    |
Db     179 VKCGPPPDISNGKHSGTEDEY--PYNHG  ISYTCDPGFELVGSPETGCIYVNKTVPVWS 234

Qy      58 KAPPICESVNKTISCSDPIVPGGFMNKGSKAPERHGDSVIFICKANFIMKGSKTVWCQAN 117
             :|: ||     | || | :  | | || : | ||    |      ::|   : || |
Db     235 SSPPICEK-    TICSQPNILNGVIVSGYKAIYFNKDSVRLACLNGTVLRGRHVIECQGN 290

Qy     118 EMWGPTALPVCESDFPL 134
           |  :::| || |  |
Db     291 GNW--SSLPDCEFDCDL 305


RESULT 9
Q29530
ID   Q29530       PRELIMINARY;      PRT;  2014 AA.
AC   Q29530;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
```

```
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94292799; PubMed=8021505;
RA   Birmingham D.J., Shen X.P., Bourcade D., Nickells M.W., Atkinson J.P.;
RT   "Primary sequence of an alternatively spliced form of CR1. Candidate
RT   for the 75,000 M(r) complement receptor expressed on chimpanzee
RT   erythrocytes.";
RL   J. Immunol. 153:691-700(1994).
DR   EMBL; L24920; AAA51438.1; -.
DR   HSSP; P08604; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER    2014   2014
SQ   SEQUENCE   2014 AA;  221281 MW;  6D6C3A74D81F1DB9 CRC64,

  Query Match             29.5%;  Score 221.5;  DB 6;  Length 2014,
  Best Local Similarity   36.8%;  Pred. No. 4.8e-16;
  Matches   50;  Conservative   17;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS   LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |         : |:||:.:.| |:.
Db    1492 ISCEPPPTISNG--DFYSNNRASFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1549

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db    1550 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 1606

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1607 QCQTNGRWGP-KLPHC 1621


RESULT 10
Q16745
ID   Q16745      PRELIMINARY;      PRT;  2039 AA.
AC   Q16745;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94065175; PubMed=8245463;
RA   Vik D.P., Wong W.W.;
RT   "Structure of the gene for the F allele of complement receptor type 1
RT   and sequence of the coding region unique to the S allele.";
RL   J. Immunol. 151:6214-6224(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Vik D.P., Wong W.W.;
RL   Submitted (JUN-1993) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L17418; AAB60694.1; -.
DR   EMBL; L17390; AAB60694.1; JOINED.
DR   EMBL; L17399; AAB60694.1; JOINED.
DR   EMBL; L17409; AAB60694.1; JOINED.
DR   EMBL; L17419; AAB60694.1; JOINED.
DR   EMBL; L17420; AAB60694.1; JOINED.
DR   EMBL; L17421; AAB60694.1; JOINED.
DR   EMBL; L17422; AAB60694.1; JOINED.
DR   EMBL; L17423; AAB60694.1; JOINED.
DR   EMBL; L17391; AAB60694.1; JOINED.
DR   EMBL; L17392; AAB60694.1; JOINED.
DR   EMBL; L17393; AAB60694.1; JOINED.
DR   EMBL; L17394; AAB60694.1; JOINED.
DR   EMBL; L17395; AAB60694.1; JOINED.
DR   EMBL; L17396; AAB60694.1; JOINED.
DR   EMBL; L17397; AAB60694.1; JOINED.
DR   EMBL; L17398; AAB60694.1; JOINED.
DR   EMBL; L17400; AAB60694.1; JOINED.
DR   EMBL; L17401; AAB60694.1; JOINED.
DR   EMBL; L17402; AAB60694.1; JOINED.
DR   EMBL; L17403; AAB60694.1; JOINED.
DR   EMBL; L17404; AAB60694.1; JOINED.
DR   EMBL; L17405; AAB60694.1; JOINED.
DR   EMBL; L17406; AAB60694.1; JOINED.
DR   EMBL; L17407; AAB60694.1; JOINED.
DR   EMBL; L17408; AAB60694.1; JOINED.
DR   EMBL; L17410; AAB60694.1; JOINED.
DR   EMBL; L17411; AAB60694.1; JOINED.
DR   EMBL; L17412; AAB60694.1; JOINED.
DR   EMBL; L17413; AAB60694.1; JOINED.
DR   EMBL; L17414; AAB60694.1; JOINED.
DR   EMBL; L17415; AAB60694.1; JOINED.
DR   EMBL; L17416; AAB60694.1; JOINED.
DR   EMBL; L17417; AAB60694.1; JOINED.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 30.
DR   SMART; SM00032; CCP; 30.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   2039 AA,  223693 MW,  B525C311C663:6635 CRC64;

  Query Match             29.5%;  Score 22 .5,  DB 4,  Length 2039;
  Best Local Similarity   36.8%;  Pred. No. 4.9e-16;
  Matches   50;  Conservative   17;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS-- ---YRLIGEKAIFCISENQV 52
           |||:| | : |    :||        |||: | |         : |:||::|:| |::
Db    1517 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 1574

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db    1575 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 1631

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    1632 QCQTNGRWGP-KLPHC 1646


RESULT 11
Q16744
ID   Q16744      PRELIMINARY;      PRT;  2439 AA.
AC   Q16744;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94065175; PubMed=8245463;
RA   Vik D.P., Wong W.W.;
RT   "Structure of the gene for the F allele of complement receptor type 1
RT   and sequence of the coding region unique to the S allele.";
RL   J. Immunol. 151:6214-6224(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RA   Vik D.P., Wong W.W.;
RL   Submitted (JUN 1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L17418; AAB60695.1; -.
DR   EMBL; L17390; AAB60695.1; JOINED.
DR   EMBL; L17399; AAB60695.1; JOINED.
DR   EMBL; L17409; AAB60695.1; JOINED.
DR   EMBL; L17419; AAB60695.1; JOINED.
DR   EMBL; L17420; AAB60695.1; JOINED.
DR   EMBL; L17421; AAB60695.1; JOINED.
DR   EMBL; L17422; AAB60695.1; JOINED.
DR   EMBL; L17423; AAB60695.1; JOINED.
DR   EMBL; L17424; AAB60695.1; JOINED.
DR   EMBL; L17425; AAB60695.1; JOINED.
DR   EMBL; L17426; AAB60695.1; JOINED.
DR   EMBL; L17427; AAB60695.1; JOINED.
DR   EMBL; L17428; AAB60695.1; JOINED.
DR   EMBL; L17429; AAB60695.1; JOINED.
DR   EMBL; L17430; AAB60695.1; JOINED.
DR   EMBL; L17391; AAB60695.1; JOINED.
DR   EMBL; L17392; AAB60695.1; JOINED.
DR   EMBL; L17393; AAB60695.1; JOINED.
DR   EMBL; L17394; AAB60695.1; JOINED.
DR   EMBL; L17395; AAB60695.1; JOINED.
DR   EMBL; L17396; AAB60695.1; JOINED.
DR   EMBL; L17397; AAB60695.1; JOINED.
DR   EMBL; L17398; AAB60695.1; JOINED.
DR   EMBL; L17400; AAB60695.1; JOINED.
DR   EMBL; L17401; AAB60695.1; JOINED.
DR   EMBL; L17402; AAB60695.1; JOINED.
DR   EMBL; L17403; AAB60695.1; JOINED.
DR   EMBL; L17404; AAB60695.1; JOINED.
DR   EMBL; L17405; AAB60695.1; JOINED.
DR   EMBL; L17406; AAB60695.1; JOINED.
DR   EMBL; L17407; AAB60695.1; JOINED.
DR   EMBL; L17408; AAB60695.1; JOINED.
DR   EMBL; L17410; AAB60695.1; JOINED.
DR   EMBL; L17411; AAB60695.1; JOINED.
DR   EMBL; L17412; AAB60695.1; JOINED.
DR   EMBL; L17413; AAB60695.1; JOINED.
DR   EMBL; L17414; AAB60695.1; JOINED.
DR   EMBL; L17415; AAB60695.1; JOINED.
DR   EMBL; L17416; AAB60695.1; JOINED.
DR   EMBL; L17417; AAB60695.1; JOINED.
DR   HSSP; P08603; 1HFI.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 37.
DR   SMART; SM00032; CCP; 37.
DR   PROSITE; PS00134; CARBOXYPEPT_ZN_2; UNKNOWN_3.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   2489 AA;  272846 MW;  CEE1B53F2B4FAF6 CRC64;

  Query Match             29.5%;  Score 221.5;  DB 4;  Length 2489;
  Best Local Similarity   36.8%;  Pred. No. 6.1e-16;
  Matches   50;  Conservative   17;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : :|||:::|:| |::
Db    1967 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 2024

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db    2025 VGVWSSPPPRCISNK---CTAPEVENAIRVPGNRSFFSLTFTVRFFCQPGFVMVGSHTV 2081

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db    2082 QCQTNGRWGP-KLPHC 2096


RESULT 12
Q9HB99
ID   Q9HB99      PRELIMINARY;      PRT;   142 AA.
AC   Q9HB99;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Miller L.H., Doumbo O., Moulds J.;
RL   Submitted (MAY 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264715; AAG14442.1; -.
DR   HSSP; P02749; 1QUB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1      1
FT   VARIANT      94     94       T -> S.
FT   NON_TER     132    132
SQ   SEQUENCE   132 AA;  14680 MW;  71A9B91B6596958B CRC64;

  Query Match             29.4%;  Score 220.5;  DB 4;  Length 142;
  Best Local Similarity   36.0%;  Pred. No. 2.7e-17;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : :|||:::|:| |::
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db      59 VGVWSSPPPRCISNK---CTAPEVENAIRVPGNRSFFTLTFTVRFFCQPGFVMVGSHTV 115

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db     116 QCQTNGRWGP-KLPHC 130


RESULT 13
Q9HB98
ID   Q9HB98      PRELIMINARY;      PRT;   142 AA.
AC   Q9HB98;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Miller L.H., Doumbo O., Moulds J.;
RL   Submitted (MAY 2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF264716; AAG14443.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     132    132
SQ   SEQUENCE   132 AA;  14666 MW;  1AC2B91B659691DA CRC64;

  Query Match             29.4%;  Score 220.5;  DB 4;  Length 142;
  Best Local Similarity   36.0%;  Pred. No. 2.7e-17;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;
```

```
Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : |:|||::|:| |::
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db      59 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIIRFRCQPGFVMVGSHTV 115

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db     116 QCQTNGRWGP-KLPHC 130


RESULT 14
Q9HC55
ID   Q9HC55       PRELIMINARY;      PRT;   132 AA.
AC   Q9HC55;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Birmingham D., Atkinson J.P.,
RA   Miller L., Doumbo O., Moulds J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF169969; AAG09139.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1      1
FT   VARIANT      85     85       G -> R.
FT   NON_TER     132    132
SQ   SEQUENCE   132 AA;  14553 MW;  1AD11B90F59691CB CRC64;

  Query Match             29.2%;  Score 219.5;  DB 4;  Length 132;
  Best Local Similarity   36.8%;  Pred. No. 3.5e-17;
  Matches   50;  Conservative   17;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:||| : |    :||        |||: | |          : |:|||::|:| |::
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db      59 VGVWSSPPPRCISTNK---CTAPEVENAIRVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 115

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db     116 QCQTNGRWGP-KLPHC 130


RESULT 15
Q9HC54
ID   Q9HC54       PRELIMINARY;      PRT;   132 AA.
AC   Q9HC54;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 (FRAGMENT).
GN   CR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zimmerman P.A., McNamara D.T., Birmingham D., Atkinson J.P.,
RA   Miller L., Doumbo O., Moulds J.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF169970; AAG09140.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 2.
DR   SMART; SM00032; CCP; 2.
KW   Receptor.
FT   NON_TER       1      1
FT   VARIANT      74     74       E -> K.
FT   VARIANT      85     85       G -> R.
FT   NON_TER     132    132
SQ   SEQUENCE   132 AA;  14554 MW;  1AD11B92F59691CB CRC64;

  Query Match             28.7%;  Score 215.5;  DB 4;  Length 132;
  Best Local Similarity   36.0%;  Pred. No. 1e-16;
  Matches   49;  Conservative   18;  Mismatches   54;  Indels   15;  Gaps    5;

Qy       2 ISCDPPPEVKNARKPYYS---LPIVPGTVLRYTCSPS------YRLIGEKAIFCISENQV 52
           |||:| | : |    :||        |||: | |          : |:|||::|:| |::
Db       1 ISCEPPPTISNG--DFYSNNRTSFHNGTVVTYQCHTGPDGEQLFELVGERSIYCTSKDDQ 58

Qy      53 HATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTV 112
             |   || | | ||   |: | |       |::: |   : | | |:  | | || ||
Db      59 VGVWSSPPPRCISTNE---CTAPEVENAIQVPGNRSFFSLTEIVRFRCQPGFVMVGSHTV 115

Qy     113 WCQANEMWGPTALPVC 128
           || |  |||  || |
Db     116 QCQTNGRWGP-KLPHC 130


RESULT 16
Q60736
ID   Q60736       PRELIMINARY;      PRT;   579 AA.
AC   Q60736;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPERM FERTILIZATION PROTEIN SP56 PRECURSOR.
GN   ZP3R OR SP56.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HSD/ICR; TISSUE=TESTIS;
RX   MEDLINE=95327956; PubMed=7604284;
RA   Bookbinder L.H., Cheng A., Bleil J.D.
RT   "Tissue- and species-specific expression of sp56, a mouse sperm
RT   fertilization protein.";
RL   Science 269:86-89(1995).
DR   EMBL; U17108; AAC52208.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:104965; Zp3r.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
KW   Signal.
FT   SIGNAL        1     32       POTENTIAL.
FT   CHAIN        33    579       SPERM FERTILIZATION PROTEIN SP56.
SQ   SEQUENCE   579 AA;  64950 MW;  C878D1F0B2C92C53 CRC64;

  Query Match             28.6%;  Score 215;  DB 1;  Length 579;
  Best Local Similarity   30.3%;  Pred. No. 6.2e-16;
  Matches   40;  Conservative   27;  Mismatches   51;  Indels   14;  Gaps    4;

Qy       3 SCDPPPEVKNAR-----KPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWD 57
           :|:||| : | :        |:    |:|: | | |:: :|  :| |   |:    :
Db     157 TCEPPPPISNGKHSGRDDDLYTF----GSVVIYNCDPTFTLIGNASIVCTVVNRIVGVWH 212
```

```
Qy      58 KAPPC.EsVNKTISCSDPTVPGGFMNKGSKAPFFRHGDSVTFTCKANFIMKGSKIVWCQAN 117
           || |.     | |  | :| | :  :  : | ::   || | | ::|| : |:||
Db     213 PHPPACQK     IVCnnnQvTTCKYtAtGFRQFYAYRDALEFCKKGFITRGSSVIRCFAN 268

Qy     118 EMWGPTALPVCE 129
           | : :: ||
Db     269 SEWFP SIPICE 279


RESULT 17
Q99JA1
ID   Q99JA1       PRELIMINARY;      PRT;   555 AA.
AC   Q99JA1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C4BF ALPHA CHAIN PRECURSOR.
GN   C4BPA.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS, AND LIVER;
RX   MEDLINE=21154058; PubMed=11254714;
RA   Nonaka M.C., Wang C., Mori T., Okada H., Nonaka M.;
RT   "Novel androgen dependent promoters direct expression of the C4b
RT   binding protein alpha chain gene in epididymis.";
RL   J. Immunol. 166:4570-4577(2001).
DR   EMBL; AB049465; BAB39737.1; -.
DR   EMBL; AB049466; BAB39738.1; -.
DR   EMBL; AB049467; BAB39739.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        14    555       POTENTIAL.
SQ   SEQUENCE   555 AA;  61628 MW;  7AFA5462AFC7B6AC CRC64;

  Query Match             28.5%;  Score 214;  DB 11;  Length 555;
  Best Local Similarity   31.6%;  Pred. No. 7.7e-16;
  Matches   42;  Conservative   25;  Mismatches   52;  Indels   14;  Gaps    4;

Qy       2 ISCDPPPEVKNAR      KPYYSLPIVPGTVLRYTCSPSYPLIGEKAIFCISENQVHATW 56
           | |  ||:: | :      : |:     |: : |:| ||| |:|  :| |   |:    |
Db     137 IKCRSPPDISNGKHSGAVEDIYIY    GSSVTYSCDPSYSLLGNPSISCTVVNKTVGVW 192

Qy      57 DKAPPICESVNKTISCSDPTIPGGFMNKGSKAPFRHGDSVTFTCKANFIMKGSKIVWCQA 116
           :|||:    | : |  |||    :  |    :   |:: |:|:  | :|||  : |:|
Db     193 SPSPIVC    KKVICkQPIVQYANIISGFGPIYTFKDTIMFSCVGKGFVLKGSSLIRCEA 248

Qy     117 EMWGPTALPVCE 129
           :  | |:  ||||
Db     249 DNNWNCSP PVCE 260


RESULT 18
Q02839
ID   Q02839       PRELIMINARY;      PRT;   363 AA.
AC   Q02839;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97454414; PubMed=9349970;
RA   Toyomura K., Fujimura T., Murakami H., Natsume T., Shigehisa T.,
RA   Inoue N., Takeda J., Kinoshita T.;
RT   "Molecular cloning of a pig homologue of membrane cofactor protein
RT   (CD46).";
RL   Int. Immunol. 9:869-876(1997).
DR   EMBL; D70897; BAA20476.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   363 AA;  39692 MW;  8A5FF329B8E03153 CRC64;

  Query Match             26.5%;  Score 199.5;  DB 4;  Length 363;
  Best Local Similarity   31.9%;  Pred. No. 2.1e-14;
  Matches   45;  Conservative   21;  Mismatches   52;  Indels   23;  Gaps    6;

Qy       1 EISCDPPPEVKNAR---- KPYYSLPIVPGTVLRYTCSPS      YRLIGEKAIFCISEN 50
           :| | || |: | :      |  :         |: ||| |      : :::| ::||| ::
Db     168 KILCKPPCEIPNGKYINSRKDVFEY    NEVVIYSCLSSYGPDEFSIVGESSLFCIGKD 223

Qy      51 QVHATWDKAPPICESVNKTISCSDPTIPGGFMNKGSKAPFRHGDSVTFTCKANFIMKGSK 110
           :    |  || |    | : |  |:|| | :  | : | :  |  | | | || : |
Db     224 E-   WSSDPPEC----KVVKCPYPVVPNGEIVSGFGSEFYYKAEVVFECNAGFTLRGRD 275

Qy     111 IVWCQANEMWGPTALPVCESD 131
           |: | |   | |  :| |  |
Db     276 TIVCGANSIWEPEMPQCIKD 296


RESULT 19
Q9R0R9
ID   Q9R0R9       PRELIMINARY;      PRT;   300 AA.
AC   Q9R0R9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SOLUBLE FORM OF MEMBRANE COFACTOR PROTEIN CD46.
GN   MCP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20094984; PubMed=10630288;
RA   Nomura M., Tsujimura A., Shida K., Matsumoto M., Matsuda Y.,
RA   Toyoshima K., Seya T.;
RT   "Membrane and secretory forms of mouse membrane cofactor protein
RT   (CD46) generated from a single gene through alternative splicing.";
RL   Immunogenetics 50:245-254(1999).
DR   EMBL; AB022600; BAA86056.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:1203290; Mcp.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   300 AA;  33658 MW;  1B57BA1655FA1E14 CRC64;

  Query Match             26.4%;  Score 198.5;  DB 11;  Length 300;
  Best Local Similarity   31.4%;  Pred. No. 2.2e-14;
  Matches   42;  Conservative   22;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVP GTVLRYTCSPS      YRLIGEKAIFCISENQVHA 54
           :|    |||:||      : :     :  :  :|| |:      : ::      |:     |
Db     170 KIYCLPPPKIKNGTHTLTHINVFKYHEAVSYSCVPTPCPNKFSLVSTSMICCAG    HN 225

Qy      55 QWDKAPPICESVNKTISCSDPTIPGGFMNKSKAPFRHGDSV T FTCKANFIMKGSKIVWC 114
           |  :|| |    | : | :|::  |  : |:  |  : : | |   | |||  | |
Db     226 TWSNSPPEC   KVVKCPNPVLQNGRLISGAGETESYQSTVMFECIQGFYMIGSSMVIC 281

Qy     115 QANEMWGPTALPVC 128
```

```
           ||  | | ::| |
Db    282 SANNSWEP-SIPKC 294


RESULT 20
O88174
ID   O88174      PRELIMINARY;      PRT;   365 AA.
AC   O88174;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (CD46).
GN   Mcp.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tsujimura A., Kitamura M., Seya T.;
RT   "Molecular cloning of a murine homologue of membrane cofactor protein
RT   (CD46): preferential expression in testicular germ cells.";
RL   Thesis (1997),
RL   Osaka Medical Center for Cancer and Cardiovascular Diseases, Japan.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98129724; PubMed=9461505;
RA   Tsujimura A., Shida K., Kitamura M., Nomura M., Takeda J., Tanaka H.,
RA   Matsumoto M., Matsumiya K., Okuyama A., Nishimune Y., Okabe M.,
RA   Seya T.;
RT   "Molecular cloning of a murine homologue of membrane cofactor protein
RT   (CD46): preferential expression in testicular germ cells.";
RL   Biochem. J. 330:163-168(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=TESTIS;
RX   MEDLINE=99016109; PubMed=9799342;
RA   Miwa T., Nonaka M., Okada N., Wakana S., Shiroishi T., Okada H.;
RT   "Molecular cloning of rat and mouse membrane cofactor protein (MCP,
RT   CD46): preferential expression in testis and close linkage between the
RT   mouse Mcp and Cr2 genes on distal chromosome 1.";
RL   Immunogenetics 48:363-371(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AB001566; BAA31859.1; -.
DR   EMBL; AB010919; BAA34810.1; -.
DR   EMBL; AK006632; BAB24682.1; -.
DR   HSSP; P10998; 1VVD.
DR   MGD; MGI:1204290; Mcp.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Membrane; Signal.
SQ   SEQUENCE   365 AA;  40881 MW;  84A4CA63E165C029 CRC64;

  Query Match             26.4%;  Score 196.5;  DB 11;  Length 365;
  Best Local Similarity   31.3%;  Pred. No. 2.7e-04;
  Matches   42;  Conservative   22;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPVKNARKPYYSLPIVP-GTVLRYTCSPS--YRLIGEKAIFCISENQVHA 54
           :| | |||::||       : :       : |:  |:      : |:|   |||
Db     170 KIYCLPPPKIKNGTHTLTDINVPKYHEAVSYSCQPTPGPDKFSLVGTSMIPCAG--HN 225

Qy      55 TWDKAEPICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFICKANFIMKGSKTVWC 114
           ||  :|| |    | : | ::::  | :  |:   | :  :| | |   | ||:||  | |
Db     226 IWSNSPPEC----KVVKCPNPVLQNGRLISGAGEIFSYQSIVMFECLQGFYMEGSSMVIC 281

Qy     115 QANEMWGPTALPVC 128
             ||      ::| |
Db     282 SANNSWEP-SIPKC 294


RESULT 21
Q89859
ID   Q89859      PRELIMINARY;      PRT;   263 AA.
AC   Q89859;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HOMOLOG OF VACCINIA VIRUS CDS C3L.
GN   D15L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BANGLADESH-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L.I., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SOMALIA-1977;
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L22579; AAA60760.1; -.
DR   EMBL; U18340; AAA69423.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28816 MW;  0465DC3D027C4220 CRC64;

  Query Match             26.3%;  Score 197.5;  DB 12;  Length 263;
  Best Local Similarity   29.9%;  Pred. No. 2.4e-14;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHAIWDKAPP 61
           :| | | : | |   |:     |:|: | |:  | ||    : |       |  ||
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSN PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFFHGDSVTFICKANFIMKGSKTVWCQANEMWG 121
           |:     : |  | :  |::: | |  : : |:| |||  : : || : |     |
Db     200 TCQ----IVKCPHPTILNGYLSSGFKRSYSYNNVDFTCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     122 PTALPVC 128
           |  || |
Db     256 P-ELPKC 261
```

```
RESULT 22
Q87044
ID   Q87044      PRELIMINARY;      PRT:   263 AA.
AC   Q87044;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   B12L PROTEIN.
GN   B12L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA 1967, SSP. MAJOR;
RX   MEDLINE=96202281; PubMed=8384129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA 1967, SSP. MAJOR;
RX   MEDLINE=95159666; PubMed=7856812;
RA   Shchelkunov S.N., Blinov V.M., Resenchuk S.M., Totmenin A.V.,
RA   Sheluma L.V., Chirikova G.B., Sandakhchiev L.S.;
RT   "Analysis of the nucleotide sequence of 53 kbp from the right terminus
RT   of the genome of variola major virus strain India 1967.";
RL   Virus Res. 34:207-236(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA 1967, SSP. MAJOR;
RX   MEDLINE=95320969; PubMed=7597802;
RA   Shchelkunov S.N., Totmenin A.V.;
RT   "Two types of deletions in orthopoxvirus genomes.";
RL   Virus Genes 9:231-245(1995).
DR   EMBL; X69198; CAA48953.1; -.
DR   PIR; C36838; C36838.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28789 MW;  5FE244F48B19D479 CRC64;

  Query Match             26.0%;  Score 195.5;  DB 12;  Length 263;
  Best Local Similarity   29.9%;  Pred. No. 2.4e-14;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGIVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  |:  : | |   |:     |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQLPPSISNGRHNGYNLEYIDGSVVIYSCNSGYSLIGNSGVICSG----GEWSN-PP 199

Qy      62 ICESVNKTISCSDPIVPGCEMNKGSKAIFKHGDSVTFICKANFIMKGSKIVWCQANEMWG 121
            |:     : |  | :  |::: | |  : : |:| ||||  : : || :  |     |
Db     200 ICQ-----IVKCPHPITINGYISSGFKRSYSYNDNVDFICKYGYKLSGSSSSTCSPDNTWQ 255

Qy     122 PTALPVC 128
              |  ||
Db     256 P-ELPKC 261


RESULT 23
Q89076
ID   Q89076      PRELIMINARY;      PRT:   263 AA.
AC   Q89076;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   B19L PROTEIN.
GN   B19L.
OS   Variola virus, and
OS   variola minor virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255, 54258;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Variola virus; STRAIN=GARCIA 1966;
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Totmenin A.V., Shchelkunov S.N., Esposito J.J.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=variola minor virus; STRAIN=GARCIA 1966;
RA   Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Esposito J.J., Sosnovtsev S.;
RT   "Analysis of the complete coding sequence of DNA of alastrim variola
RT   minor virus strain Garcia 1966.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U18338; AAA69358.1; -.
DR   EMBL; Y16780; CAB54612.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28812 MW;  B565ID91627C423D CRC64;

  Query Match             26.0%;  Score 195.5;  DB 12;  Length 263;
  Best Local Similarity   29.9%;  Pred. No. 4.1e-14;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGIVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  |:  : | |   |:     |:|: |:|:  | |||   : |         |   ||
Db     146 VKCQLPPSISNGRHNGYNDEYIDGSVVIYSCNSGYSLIGNSGVICSG----GEWSN-PP 199

Qy      62 ICESVNETISCSDPIVPGCEMNKGSKAIFKHGDSVTFICKANFIMKGSKIVWCQANEMWG 121
            |:     : |  | :  |::: | |  : : |:| ||||  : : || :  |     |
Db     200 ICQ-----IVKCPHPITINGYISSGFKRSYSYNDNVDFICKYGYKLSGSSSSTCSPDNTWQ 255

Qy     122 PTALPVC 128
              |  ||
Db     256 P-ELPKC 261


RESULT 24
Q89061
ID   Q89061      PRELIMINARY;      PRT:   263 AA.
AC   Q89061;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   D15L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CONGO 1965;
RX   MEDLINE=96295428; PubMed=8661439;
RA   Massung R.F., Loparev V.N., Knight J.C., Totmenin A.V.,
RA   Chizhikov V.E., Parsons J.M., Safronov P.F., Gutorov V.V.,
RA   Shchelkunov S.N., Esposito J.J.;
RT   "Terminal region sequence variations in variola virus DNA.";
RL   Virology 221:291-300(1996).
DR   EMBL; U18337; AAA69317.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   263 AA;  28843 MW;  046508766EAC4220 CRC64;

  Query Match             26.0%;  Score 195.5;  DB 12;  Length 263;
```

```
  Best Local Similarity   29.9%;  Pred. No. 4.1e-14;
  Matches   38;  Conservative   20;  Mismatches   58;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |:     |:|: |:|:  | |||   : |          |   ||
Db     146 VKCQLPPSISNGRHNGYNDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSN-PP 199

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
            |:     : |  | :  |::: | |  : : |:| |||  : : || :  |     |
Db     200 TCQ---IVKCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWQ 255

Qy     122 PTALPVC 128
           |  || |
Db     256 PELPKC 261


RESULT 25
PB7616
ID   PB7616       PRELIMINARY;      PRT;   259 AA
AC   PB7616;
DT   01-MAY-1997 (TrEMBLrel. 04, Created)
DT   01-MAY-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   41KBP FRAGMENT FROM LEFT END OF GENOME.
GN   D17L OR C17L.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=97068532; PubMed=8963248;
RA   Safronov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,
RA   Shchelkunov S.N., Sandakhchiev L.S.;
RT   "Genes of a circle of hosts for the cowpox virus.",
RL   Dokl. Akad. Nauk 349:829-833(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GRI-90;
RX   MEDLINE=98229462; PubMed=9568042;
RA   Shchelkunov S.N., Safronov P.F., Totmenin A.V., Petrov N.A.,
RA   Ryazankina O.I., Gutorov V.V., Kotwal G.J.;
RT   "The genomic sequence analysis of the left and right species-specific
RT   terminal region of a cowpox virus strain reveals unique sequences and
RT   a cluster of intact ORFs for immunomodulatory and host range
RT   proteins.";
RL   Virology 243:432-460(1998).
DR   EMBL; X94355; CAA64102.1; -.
DR   EMBL; Y11842; CAA72567.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   259 AA;  28193 MW;  9D1AAEF6893B859A CRC64;

  Query Match             25.7%;  Score 193.5;  DB 12;  Length 259;
  Best Local Similarity   29.3%;  Pred. No. 6.8e-14;
  Matches   87;  Conservative   19;  Mismatches   60;  Indels   11;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAPP 61
           : |  || : | |   |      |:|: |:|:  | |||   : |          |   ||
Db     142 VKCQSPPSISNGRHNGYFEDFYTDGSVVTYSCNSGYSLIGNSGVLCSG-----GEWSD-PP 195

Qy      62 ICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQANEMWG 121
            |:     : |  | :  |::: | |  : : |:| | ||  : : || :  |     |
Db     196 TCQ----IVKCPHPTISNGYLSSGFKRSYSYNDNVDFKCKYGYKLSGSSSSTCSPGNTWQ 251

Qy     122 PTALPVC 128
           |  || |
Db     252 PELPKC 257


RESULT 26
Q15429
ID   Q15429       PRELIMINARY;      PRT;   349 AA
AC   Q15429;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CD46.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Hara T., Suzuki Y., Nagasawa S., Seya T.;
RT   "cDNA cloning and characterization of human sperm CD46.";
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D84105; BAA12224.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   349 AA;  39325 MW;  8EFCEDA30D3C318E CRC64;

  Query Match             25.1%;  Score 188.5;  DB 4;  Length 349;
  Best Local Similarity   30.6%;  Pred. No. 3.5e-13;
  Matches   41;  Conservative   23;  Mismatches   55;  Indels   15;  Gaps    6;

Qy       1 EISCDPPLEVKNARKPYYSLPIVPG-TVLRYTCSPS     -YRLIGEKAIFCISENQVHA 54
           :: |  ||::|| :  :  : :      : |:| |:     : |||| |:|  :| |
Db     159 KVLCTPPGKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESTIYCGDNSV   215

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           | :| | |    | : |  |:|  |    |    | :  :| | |   | : || |: |
Db     216 -WSKAAPRC----KVVKCRFPVVENGKQISGFGKKFYYKAIVMFECDKGFYLDGSDTIVC 270

Qy     115 QANEMWGPTALPVC 128
           :|    |  :| |
Db     271 DSNSTWDPPVPKC 283


RESULT 27
Q9WRU2
ID   Q9WRU2       PRELIMINARY;      PRT;   645 AA.
AC   Q9WRU2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT BINDING PROTEIN.
OS   Macaca mulatta rhadinovirus 17577.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=85534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99174001; PubMed=10074154;
RA   Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RT   "Sequence and genomic analysis of a rhesus macaque rhadinovirus with
RT   similarity to Kaposi's sarcoma-associated herpesvirus/human
RT   herpesvirus 8.";
RL   J. Virol. 73:3040-3053(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Searles R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083501; AAD21332.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 8.
DR   SMART; SM00032; CCP; 8.
SQ   SEQUENCE   645 AA;  71526 MW;  93LBEF5ABF61EB2 CRC64;
```

```
  Query Match             24.7%;  Score 185.5;  DB 12;  Length 645;
  Best Local Similarity   31.4%;  Pred. No. 1.6e-12;
  Matches   44;  Conservative   25;  Mismatches   46;  Indels   25;  Gaps    7;

Qy       4 CDPPEEVKNAR    KPYYSLPIVPGIVLRYTCSPSYRLIGEKAIPCISENQVHAIWDKA 59
           | |||:| | :    | ::          : :::|  : |:|::   |||       |:|:|
Db     449 CKPPPQLINCKYHPVKDFYQYI    DIVIFSCNRDPSLVGDEMTICISN      IWNKP 499

Qy      60 PPICESVNKTISCSDPIVPGGFMNKGSKAPERHGDSVIFICKANEIMKGSKIVWCQANEM 119
           | ||     |:|| | :  |  :: :|| |||  |:  ||: ||:   |: :
Db     500 PPRCPEQ    IPCSAPNIAHGKILDGSSSVYKYGQSVFIGCFPGFTLIGSEISICK-DSS 554

Qy     120 WGPTALPVC       ESLEP 133
           : |  || |         | |
Db     555 WDP PLPICVPAVSMPSLIP 575


RESULT 29
Q15912
ID   Q15912       PRELIMINARY;      PRT;   285 AA.
AC   Q15912;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN   CD46.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97306679; PubMed=9224509;
RA   Hsu E.C., Doria R.E., Saranqi F., Marcil A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
DR   EMBL; U82015; AAB66815.1; -.
DR   HSSP; P10998; 1VVL.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
FT   NON_TER      285      285
SQ   SEQUENCE   285 AA;  31046 MW;  40AA8FA700299444 CRC64;

  Query Match             24.5%;  Score 184.5;  DB 6;  Length 285;
  Best Local Similarity   29.9%;  Pred. No. 7.9e-13;
  Matches   40;  Conservative   24;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKFYYSLPIVPG TVLRYTCSPS      YRLIGEKAIPCISENQVHA 54
           :| | |||::|| |   : : :       : |:| |:      : |||  |:| :    ::
Db     159 KTLCIPPPKIKNGKHTFSEVAVFEYYIDAVTYSCDPAPGPEPFSLIGESSMIYCGN ---NS 214

Qy      55 IWDKAPPICESVNKTIs SCSDPIVPGGFMNKGSKAPERHGDSVIFICKANEIMKGSKIVWC 114
           ||   :  |    | : | :: |    |    | :  :|| | |    : : || : |
Db     215 IWSHAAPEC    KVVKCPFPVVENGKQISGFGKEFYYKAIVMEFCDPGYYLNGSDEIVC 270

Qy     115 QANEMWGPTALPVC 128
           :: | | |  :| |
Db     271 ESNSTWDP PVPKC 285


RESULT 29
Q99Y21
ID   Q99Y21       PRELIMINARY;      PRT;  1001 AA.
AC   Q99Y21;
DT   01-NOV-1999 (TrEMBLrel. 1, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0927 PROTEIN (FRAGMENT).
GN   KIAA0927.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
CC   -!- SIMILARITY: CONTAINS 3 CUB DOMAINS.
DR   EMBL; AB023144; BAA76771.2; -.
DR   HSSP; P29392; 1SFP.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
DR   SMART; SM00042; CUB; 3.
DR   PROSITE; PS01180; CUB; 3.
FT   NON_TER      1      1
SQ   SEQUENCE   1001 AA;  108690 MW;  3F0B9EC8812F9491 CRC64;

  Query Match             24.5%;  Score 184.5;  DB 4;  Length 1001;
  Best Local Similarity   30.8%;  Pred. No. 3.4e-12;
  Matches   40;  Conservative   21;  Mismatches   66;  Indels    9;  Gaps    3;

Qy       3 SCDPPFVFNAPFPYYSLPIVPGTVLRYTCSPSYRLIGEKAIP CISENQVHAIWDKAPPI 62
           ||   |:|:|  |      :| |  : | | | | ::|   : |     |   :|   ||
Db     796 SCSDIPEIQNGWKTTSHTELVRGARITYQCDPGYDIVGSDTLTC    QWDLSWSSDPPF 851

Qy      63 CESVNKTISCSDPIVPGGFMNKGSKAPERHGDSVIFICKANEIMKGSKIVWC  QANEM 119
           ||   :::::| :| :|       |  :  |:|:||:|   | : .  |: '   | :
Db     852 CEKTEESLACDNPGLPENGYQILYKRLYLRGESLTFMCYEGFEIMGEVIIRCIISQPSHW 911

Qy     120 WGPTALPVCE 129
             |:  :|||:
Db     912 NGP--LPVCK 919


RESULT 30
Q29531
ID   Q29531       PRELIMINARY;      PRT;   661 AA.
AC   Q29531;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR 1 PRECURSOR (FRAGMENT).
GN   CR1.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94292799; PubMed=8021505;
RA   Birmingham D.J., Shen X.P., Bourcade D., Nickells M.W., Atkinson J.P.;
RT   "Primary sequence of an alternatively spliced form of CR1. Candidate
RT   for the 75,000 M(r) complement receptor expressed on chimpanzee
RT   erythrocytes.";
RL   J. Immunol. 154:691-700(1994).
DR   EMBL; L24921; AAA51439.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR001424; SOD_CU_ZN.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 9.
DR   PRINTS; PR00344; SELECTIN.
```

```
DR   SMART; SM00032; CCP; 9.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
KW   Signal; Receptor.
FT   NON_TER       1      1
FT   SIGNAL       -1     16       POTENTIAL.
FT   CHAIN        17    >661      COMPLEMENT RECEPTOR 1.
FT   NON_TER     661    661
SQ   SEQUENCE   661 AA;  72966 MW;  4F0DF1532664D3F0 CRC64;

  Query Match             24.4%;  Score 183.5;  DB 6;  Length 661;
  Best Local Similarity   33.3%;  Pred. No. 2.7e-12;
  Matches   45;  Conservative   21;  Mismatches   48,  Indels   21,  Gaps    7,

Qy       8 SCDPPPEVKNARK-PYYSLPIVPGIVLRYTCSPSYRLIGEKAIFCISENQVHATWDKAP 66
           :|  ||:::|      : ||  :||  : | | | | |:|:  |||  :      | :
Db     469 ACPHPPKIQNGHDIGGHVSL-YLPGMTISYICDPGYLLVGKGFIFCTDQ----GIWSQLD 523

Qy      67 PICESVNKTISCSDPIVPGGFMNKGS-----KAPFRHGDSVTFTCKANFTMKGSKTVWCQ 121
             |: ||    || |:    ||| |      | : :|| ||  |:  :|:::||   ||
Db     524 HYCKEVN----CSFPL----FMNGISKELEMKKVYHYGDYVTLKCEDGYTLEGSPWSQCQ 575

Qy     122 ANEMWGPTALPVCES 136
           |:: | |  |  | |
Db     576 ADDRWDP-PLAKCIS 589


RESULT 31
Q28770
ID   Q28770      PRELIMINARY;      PRT;   315 AA.
AC   Q28770;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT RECEPTOR (FRAGMENT).
OS   Papio cynocephalus (Yellow baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RA   Birmingham D.J., Logar C.M., Shen X.-P., Chen W.;
RT   "The baboon erythrocyte complement receptor is a glycophosphatidyl
RT   inositol-linked protein encoded by a homologue of the human CR1-like
RT   genetic element.";
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L77978; AAA99005.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 5.
DR   SMART; SM00032; CCP; 5.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     315    315
SQ   SEQUENCE   315 AA;  35172 MW;  305B314AE752824F CRC64;

  Query Match             24.4%;  Score 182.5;  DB 6;  Length 315;
  Best Local Similarity   32.4%;  Pred. No. 1.5e-12;
  Matches   45;  Conservative   18;  Mismatches   57;  Indels   19;  Gaps    5;

Qy       1 EISCDPPPEVKN-    -ARKPYYSLPIVPGIVLRYTCS------PSYRLIGEKAIFCISE 49
           :|||  || : |     | | |:      |:|: | |:          : |:|| :|:| |
Db     122 KISCGLPPTIDNGDFFSANKEYFHY----GSVVTYRCNLGSGGRKLFELVGEPSIYCTSN 177

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
             |   | |   ||   |: | |  | :   ::: |   : | | |:  | |||
Db     178 EDQVGIWSGPAPQCIIPNK---CTPPNVENGILVSVNRSLFSLNEVVEFRCQPGFVMKGP 234

Qy     110 KTVWCQANEMWGPTALPVC 128
           : | |||   | |  || |
Db     235 RRVQCQALNKWEP-ELPSC 252


RESULT 32
O19126
ID   O19126      PRELIMINARY;      PRT;   285 AA.
AC   O19126;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN   CD46.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97366679; PubMed=9223509;
RA   Hsu E.C., Iorio R.E., Sarangi F., Marc I A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
DR   EMBL; U87921; AAB66821.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
FT   NON_TER     285    285
SQ   SEQUENCE   285 AA;  31998 MW;  629ED3FED53FEE4 CRC64;

  Query Match             24.1%;  Score 181.5;  DB 6;  Length 285;
  Best Local Similarity   29.1%;  Pred. No. 1.7e-12;
  Matches   39;  Conservative   25;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-IVL-YTCSPS----YRLIGEKAIFCISENQVHA 54
           :| |  ||:::|| :  :  : :      : |:  |:     : |||| |:| :    ::
Db     159 KILCTPPPKIKNGKHTFSEVEVFEYLDAVIYSCDPAPGPDPFSLIGESMIYCGN-    NS 214

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           ||  | | |    | : |  |:|  |    |    | :  :| | |   : : ||  : |
Db     215 TWSHAAPEC----KVVKCRFPVVENGKQISGFGKKFYYKATVMFECDKGYYINGSDKIVC 270

Qy     115 QANEMWGPTALPVC 128
           ::|    |  :| |
Db     271 ESNSTWDP-PVPKC 283


RESULT 33
O19127
ID   O19127      PRELIMINARY;      PRT;   285 AA.
AC   O19127;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN   CD46.
OS   Macaca mulatta (Rhesus macaque).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9544;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97366679; PubMed=9223509;
RA   Hsu E.C., Iorio R.E., Sarangi F., Marcil A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
DR   EMBL; U87922; AAB66822.1; -.
```

```
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
FT   NON_TER      285    285
SQ   SEQUENCE   285 AA;  31998 MW;  6290D3FEED53FEE4 CRC64;

  Query Match             24.1%;  Score 181.5;  DB 6;  Length 285;
  Best Local Similarity   29.1%;  Pred. No. 1.7e-12;
  Matches   39;  Conservative   25;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARRPYYSLPIVPC IVLEYTCSPS      YRLIGEKAIFCISENQVHA 54
           | | ||||:|| :    : :       : |:| |:      : |||| |:| :    ::
Db     159 KILCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN    NS 214

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           || : | |    | : |  |:|  |    |     | :  :' | |   : :|  : |
Db     215 TWSHAAPSC    KVVKCKFPVVENGKQISGFGKKFYYKAIVMEFDCKGYYSNCSTKIVC 270

Qy     115 QANEMWGPTALPVC 128
           ::|  | |  :| |
Db     271 ESNSTWDP-PVPKC 283


RESULT 34
P79138
ID   P79138       PRELIMINARY;      PRT;   369 AA.
AC   P79138;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (MCP, CD46).
GN   CD46.
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97149841; PubMed=8996635;
RA   Murakami Y., Seya T., Kurita M., Nagasawa S.;
RT   "Molecular cloning of a complementary DNA for a membrane cofactor
RT   protein (MCP, CD46)/measles virus receptor on Vero cells and its
RT   functional characterization.";
RL   Biol. Pharm. Bull. 19:1541-1545(1996).
RN   [2]
RP   SEQUENCE OF 1-285 FROM N.A.
RX   MEDLINE=97366679; PubMed=9223509;
RA   Hsu E.C., Dorig R.E., Sarangi F., Marcil A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
RN   [3]
RP   SEQUENCE OF 1-285 FROM N.A.
RA   Richardson C.D., Hsu E.C.H., Sarangi F., Dorig R.E.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D63811; BAA09881.1; -.
DR   EMBL; U87920; AAB66820.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Membrane.
SQ   SEQUENCE   369 AA;  41670 MW;  26E16D15AF768688 CRC64;

  Query Match             24.1%;  Score 181.5;  DB 6;  Length 369;
  Best Local Similarity   29.1%;  Pred. No. 2.3e-12;
  Matches   39;  Conservative   25;  Mismatches   55;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARRPYYSLPIVPC IVLEYTCSPS      YRLIGEKAIFCISENQVHA 54
           | | ||||:|| :    : :       : |:| |:      : |||| |:| :    ::
Db     159 KILCTPPPKIKNGKHTFSEVEVFEYLDAVTYSCDPAPGPDPFSLIGESMIYCGN    NS 214

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           || : | |    | : |  |:|  |    |     | :  :' | |   : :|  : |
Db     215 TWSHAAPSC    KVVKCKFPVVENGKQISGFGKKFYYKAIVMEFDCKGYYSNCSTKIVC 270

Qy     115 QANEMWGPTALPVC 128
           ::|  | |  :| |
Db     271 ESNSTWDP-PVPKC 283


RESULT 35
Q90QV2
ID   Q90QV2       PRELIMINARY;      PRT;   559 AA.
AC   Q90QV2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC 2001 (TrEMBLrel. 19, Last annotation update)
DE   CR1 C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89010527; PubMed=2971757;
RA   Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;
RT   "Identification of an alternative polyadenylation site in the human
RT   C3b/C4b receptor (complement receptor type 1) transcriptional unit and
RT   prediction of a secreted form of complement receptor type 1.";
RL   J. Exp. Med. 168:1255-1270(1988).
DR   EMBL; X14362; CAA32541.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR002396; Selectin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR000834; Zn_carbopept.
DR   Pfam; PF00084; sushi; 8.
DR   PRINTS; PR00343; SELECTIN.
DR   SMART; SM00032; CCP; 8.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2, UNKNOWN 1.
KW   Signal; Receptor.
FT   NON_TER       1      1
FT   SIGNAL        <1     16       POTENTIAL.
FT   CHAIN         17     559      POTENTIAL.
SQ   SEQUENCE   559 AA;  61424 MW;  DBFFE069CA179D75 CRC64;

  Query Match             23.7%;  Score 178.5;  DB 4;  Length 559;
  Best Local Similarity   30.9%;  Pred. No. 8.2e-12;
  Matches   43;  Conservative   20;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPEVKNA---  --RRPYYSLPIVPCIVLEYTCSPS      YRLIGEKAIFCISE 49
           | |  || : |       |: ::      |:|: | |:       : |:|| :|:| |
Db     139 IPCGLPPTINGDFISTNRENFHY     GSVVIYRCNPGSGGRKVFELVGEPSIYCISN 193

Qy      50 NQVHAIWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
           :    |   | |  ||   |: | |  | :   ::: |   : | | |:  | |||
Db     194 DDQVGIWSGPAPQCIIPNK   CTPPNVENGILVSDNRSLFSLNEVVEFRCQPGFVMKGP 250

Qy     110 KTVWCQANEMWGPTALPVC 128
           : | |||   | |  || |
Db     251 RRVECQALNKWEPELPSC 268


RESULT 36
Q9DEC0
ID   Q9DEC0       PRELIMINARY;      PRT;   451 AA.
AC   Q9DEC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC 2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT REGULATORY MEMBRANE PROTEIN.
GN   CREMP.
```

```
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20521924; PubMed=11123420;
RA   Inoue N., Fukui A., Nomura M., Matsumoto M., Toyoshima K., Seya T.;
RT   "A novel chicken membrane associated complement regulatory protein:
RT   Molecular cloning and functional characterization.";
RL   J. Immunol. 166:424-431(2001).
DR   EMBL; AB035592; BAB16878.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
SQ   SEQUENCE   451 AA;  48723 MW;  4FDE12139A502837 CRC64;

  Query Match             23.5%;  Score 177;  DB 13;  Length 451;
  Best Local Similarity   28.4%;  Pred. No. 9.5e-12;
  Matches   38;  Conservative   21;  Mismatches   63;  Indels   12;  Gaps    4;

Qy       2 ISCDPPPEVKNARKPYYSLPIVPGTVLRYTCSP------SYRLIGEKAIFCISENQVHAT 55
           : | ||| : | |    :  | | : |:|        : |:|  :|||  :   :
Db     148 VPCKPPPSIANGRYT-EAANVYVYQTTVTYSCDDVRTGENPESLVGSPSIFCTVDENSNGV 206

Qy      56 WDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWCQ 115
           |   || |    | : | :| | |    |  : : :| || | |  :: : |   : |
Db     207 WSGPPPQC----KVVICDNPQVENGRKASGFASQYIYGSSVRFECDPDYVLLGMDVISCT 262

Qy     116 ANEMWGPTALPVCE 129
             |  | | :|| |:
Db     263 ENGIWYP-SLPTCK 275


RESULT 37
O19120
ID   O19120      PRELIMINARY;      PRT;   222 AA.
AC   O19120;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN   CD46.
OS   Aotus trivirgatus (Night monkey) (Douroucouli).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=9505;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97366679; PubMed=9223509;
RA   Hsu E.C., Dorig R.E., Sarangi F., Marcil A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
DR   EMBL; U87914; AAB66814.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
FT   NON_TER     222     222
SQ   SEQUENCE   222 AA;  24469 MW;  D92AC47BE0DB0F0D CRC64;

  Query Match             23.3%;  Score 175.5;  DB 6;  Length 222;
  Best Local Similarity   28.4%;  Pred. No. 6.2e-12;
  Matches   38;  Conservative   22;  Mismatches   59;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :| | ||| :|| :  :  : :       : |:| |:     : |:||  |:|    :
Db      96 KILCKPPPTIKNGKHTFSEVLVFEYLDAVTYSCDPAPGPDPFSLVGESTIYC----RDSL 151

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           |   |  |    | : |  |:: |    :  |  : :| | |    | : |  |  : || |: |
Db     152 GWSGDPPEC----KVVKCRFPVIENGKQTAGFG-KFYYNAAVMFECYEGFHIGSDTLVC 207

Qy     115 QANEMWGPTALPVC 128
             :|  | |  :| |
Db     208 NSNSTWDP-PVPKC 220


RESULT 38
O19125
ID   O19125      PRELIMINARY;      PRT;   222 AA.
AC   O19125;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN (FRAGMENT).
GN   CD46.
OS   Saimiri sciureus (Common squirrel monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX   NCBI_TaxID=9521;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97366679; PubMed=9223509;
RA   Hsu E.C., Dorig R.E., Sarangi F., Marcil A., Iorio C.,
RA   Richardson C.D.;
RT   "Artificial mutations and natural variations in the CD46 molecules
RT   from human and monkey cells define regions important for measles virus
RT   binding.";
RL   J. Virol. 71:6144-6154(1997).
DR   EMBL; U87913; AAB66819.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 3.
DR   SMART; SM00032; CCP; 3.
FT   NON_TER     222     222
SQ   SEQUENCE   222 AA;  24517 MW;  75DE1D756B863C78 CRC64;

  Query Match             23.3%;  Score 175.5;  DB 6;  Length 222;
  Best Local Similarity   28.4%;  Pred. No. 6.2e-12;
  Matches   38;  Conservative   22;  Mismatches   59;  Indels   15;  Gaps    5;

Qy       1 EISCDPPPEVKNARKPYYSLPIVPG-TVLRYTCSPS-----YRLIGEKAIFCISENQVHA 54
           :| | ||| :|| :  :  : :       : |:| |:     : |:||  |:|    :
Db      96 KILCKPPPTIKNGKHTFSEVDVFEYLDAVTYSCDPAPGPDPFSLVGESTIYC----RDSL 151

Qy      55 TWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGSKTVWC 114
           |   |  |    | : |  |:: |    :  |  : :| | |    | : |  |  : || |: |
Db     152 GWSGDPPEC----KVVKCRFPVIENGKQTAGFG-KFYYNAAVMFECYEGFHIGSDTLVC 207

Qy     115 QANEMWGPTALPVC 128
             :|  | |  :| |
Db     208 NSNSTWDP-PVPKC 220


RESULT 39
O62834
ID   O62834      PRELIMINARY;      PRT;   345 AA.
AC   O62834;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN CD46.
OS   Saguinus oedipus (Cotton-top tamarin).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.
OX   NCBI_TaxID=9490;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B95-8;
RX   MEDLINE=98161817; PubMed=9494106;
```

```
RA   Murakami Y., Seya T., Kurita M., Fukui A., Ueda S., Nagasawa S.;
RT   "Molecular cloning of membrane cofactor protein (MCP; CD46) on B95a
RT   cell, an Epstein-Barr virus transformed marmoset B cell line: B95a MCP
RT   is susceptible to infection by the CAM, but not the Nagahata strain of
RT   the measles virus.";
RL   Biochem. J. 330:1351-1359(1998).
DR   EMBL; D63848; BAA25596.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
SQ   SEQUENCE   336 AA;  36238 MW;  548950D54657024F CRC64;

  Query Match             23.8%;  Score 175.5;  DB 6;  Length 336;
  Best Local Similarity   28.8%;  Pred. No. 1e-11;
  Matches   40;  Conservative   24;  Mismatches   61;  Indels   15;  Gaps    6;

Qy       1 EISCDPPPEVKNARKEPYYSIEIVVG IVLFYLCSPS     YRLIGEKAIFCISENQVHA 54
           :: | .|||:: | :  : : :      : |:| |:     : |||  |:|  :| |
Db      95 KITSKEPPPKINNGKHIESIVDVFEYLDAVIYSCDPAPGPDDPFSLIGESIIYC RDNSV- 152

Qy      55 IWDKAIPICESVNKIISCSDPIVPGGEMNKGSKAPFRHGDSVIFTCKANFTMKGSKIVWC 114
           | | |  |  :  | |:  |:  |   |   |  : :| | |  : || |:| |
Db     153  WSGDAPEC    KVVKCSEPVIENGKQIAGFGKKFYYKATVIFECDEGFHIIGSDEIVC 207

Qy     115 QANEMWGPTALPVCESEPE 133
           |  | | |  .|    : |
Db     208 NSNSTWDP PVPKCVKELE 225


RESULT 40
Q920M4
ID   Q920M4       PRELIMINARY;       PRT;   355 AA.
AC   Q920M4;
DT   1 MAY 1999 (TrEMBLrel. 10, Created)
DT   1 MAY 1999 (TrEMBLrel. 10, Last sequence update)
DT   31 JUN 2001 (TrEMBLrel. 17, Last annotation update)
DE   MEMBRANE COFACTOR PROTEIN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE DAWLEY; TISSUE=TESTIS;
RX   MEDLINE=99016109; PubMed=9799432;
RA   Miwa T., Nonaka M., Okada N., Wakana S., Shiroishi T., Okada H.;
RT   "Molecular cloning of rat and mouse membrane cofactor protein (MCP,
RT   CD46): preferential expression in testis and close linkage between the
RT   mouse Mcp and Cr2 genes on distal chromosome 1.";
RL   Immunogenetics 48:463-471(1998).
DR   EMBL; AB015920; BAA34811.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 4.
DR   SMART; SM00032; CCP; 4.
KW   Signal.
FT   SIGNAL        1     41       POTENTIAL.
FT   CHAIN        42    355       MEMBRANE COFACTOR PROTEIN.
SQ   SEQUENCE   355 AA;  39267 MW;  01DFB7033ACB41B4 CRC64;

  Query Match             23.1%;  Score 174;  DB 11;  Length 355;
  Best Local Similarity   25.6%;  Pred. No. 1.6e-11;
  Matches   41;  Conservative   24;  Mismatches   60;  Indels   30;  Gaps    6;

Qy       1 EISCDPPPEVKNARKEPYYSIEIVVG GIVLFYLCSPS     YRLIGEKAIFCISENQVHA 54
           :: | :||::||  :  : :      : |:| |:     : |:|   ::|     |
Db     168 KVYCLPPPKIENGIHTESDIKEYGYQEAVIYSCDPNPGPDKFSLVGESMLEYAC    HN 223

Qy      55 IWDKAPPICESVNKIISCSDPIVPGGEMNKGSKAPFRHGDSVIFTCKANFTMKGSKIVWC 114
           ||  || |    | : |  |:  |     ::  | :   | | |   | |:||  | |
Db     224 IWSSDQPEC    KVVKCPEPVEQNGRQISKIEKKFSYQALVLFQCLEGFYMEGSSMVVC 279

Qy     115 QANEMWGPTALPVC      ESDEP 133
           |  | | |  .|          |:|
Db     280 CAKSSWEP SIPQCLEGPEIHSTEPPVYSESGYP 312


RESULT 41
Q9ES77
ID   Q9ES77       PRELIMINARY;       PRT;  4567 AA.
AC   Q9ES77;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POLYDOM PROTEIN PRECURSOR.
GN   POLYDOM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/HENSIC;
RX   MEDLINE=20517255; PubMed=11062057;
RA   Gilgès D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA   Romeo P.-H., Vigon I.;
RT   "Polydom : a secreted protein with pentraxin, complement control
RT   protein, epidermal growth factor and von Willebrand factor A
RT   domains.";
RL   Biochem. J. 352:49-59(2000).
DR   EMBL; AF206329; AAG42160.1; -.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1928849; Polydom.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF_like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_1.
DR   InterPro; IPR003410; HYR.
DR   InterPro; IPR001759; Pentaxin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR002035; vWFA.
DR   Pfam; PF00008; EGF; 10.
DR   Pfam; PF02494; HYR; 2.
DR   Pfam; PF00084; sushi; 33.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00895; PENTAXIN.
DR   PRINTS; PR00453; VWFADOMAIN.
DR   ProDom; PD002153; Pentaxin; 1.
DR   SMART; SM00032; CCP; 34.
DR   SMART; SM00181; EGF; 15.
DR   SMART; SM00179; EGF_CA; 9.
DR   SMART; SM00001; EGF_like; 3.
DR   SMART; SM00159; PTX; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Signal.
FT   SIGNAL        1     17       POTENTIAL.
SQ   SEQUENCE  4567 AA;  487491 MW;  8FBA8276E1229E5 CRC64;

  Query Match             23.0%;  Score 173;  DB 11;  Length 4567;
  Best Local Similarity   29.4%;  Pred. No. 2.9e-10;
  Matches   35;  Conservative   23;  Mismatches   49;  Indels   12;  Gaps    4;

Qy       2 ISCDPPPEVKNARKEPYYSIEIVVGTVLFYTCSPSYRLIGEKAIFCISENQVEAIWSEAPP 61
           : |  | :: |  |   :  ::  ::|   | | |   :   :      :|:|||
Db    1902 VKCSQPEDINNGKIEYTLSGLIYLSIASYSCENGYSLQGPSLIECVAS  SSWQKAPP 1955

Qy      62 ICESVNKTISCSDPIVPGGEMNKGSKAPFRHGDSVIFTCKANFTMKGSKIVWCQANEMW 120
           |:     :|| :| :    :  ||   |  |::| :|||  :|  | |  |: |||| |
Db    1956 SCQ---LVSCGEPPIVKDAVIIGSN--EIEGNIVAYICKEGYILAGPDIIVCQANGKW 2008
```

```
RESULT 42
O45520
ID   O45520      PRELIMINARY;      PRT:   417 AA.
AC   O45520;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   5I2 ANTIGEN (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RT   "Novel isoform of rat complement regulatory protein, rat Crry.";
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D42116; BAA22549.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
FT   NON_TER      417     417
SQ   SEQUENCE   417 AA;  45951 MW;  042D3B68D9E89BDB CRC64;

  Query Match             22.9%;  Score 172.5;  DB 11;  Length 417;
  Best Local Similarity   30.2%;  Pred. No. 2.8e-11;
  Matches   42;  Conservative   21;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |: || : |       |: ::      | |: | |:           : :|| :| | |
Db      21 IPCEIPPSIPNGDFFSPNREDFHY-----GMVVTYQCNTDARGKKLFNLVGEPSIHCTSI 75

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
           :     |   || |  :||   |: | |    :   :|: |   | | | |:  | |||
Db      76 DGQVGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVFFRCQDGFMMKGD 132

Qy     110 KTVWCQANEMWGPTALPVC 128
           :|:|::   | |  || |
Db     133 SSVYCRSLNRWEP-QLPSC 150


RESULT 43
Q64612
ID   Q64612      PRELIMINARY;      PRT:   497 AA.
AC   Q64612;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   5I2 ANTIGEN PRECURSOR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=94161746; PubMed=8117286;
RA   Sakurada C., Seno H., Dohi N., Takizawa H., Nonaka M., Okada N.,
RA   Okada H.;
RT   "Molecular cloning of the rat complement regulatory protein, 5I2
RT   antigen.";
RL   Biochem. Biophys. Res. Commun. 198:819-826(1994).
DR   EMBL; D42114; BAA07698.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 6.
DR   SMART; SM00032; CCP; 6.
KW   Signal.
FT   SIGNAL        1     46       POTENTIAL.
FT   CHAIN        47    497       5I2 ANTIGEN.
SQ   SEQUENCE   497 AA;  54786 MW;  177AC11EE0F1AD1C CRC64;

  Query Match             22.9%;  Score 172.5;  DB 11;  Length 497;
  Best Local Similarity   30.2%;  Pred. No. 3.4e-11;
  Matches   42;  Conservative   21;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |: || : |       |: ::      | |: | |:           : :|| :| | |
Db     159 IPCEIPPSIPNGDFFSPNREDFHY-----GMVVTYQCNTDARGKKLFNLVGEPSIHCTSI 213

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
           :     |   || |  :||   |: | |    :   :|: |   | | | |:  | |||
Db     214 DGQVGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVFFRCQDGFMMKGD 270

Qy     110 KTVWCQANEMWGPTALPVC 128
           :|:|::   | |  || |
Db     271 SSVYCRSLNRWEP-QLPSC 288


RESULT 44
Q63135
ID   Q63135      PRELIMINARY;      PRT:   559 AA.
AC   Q63135;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   COMPLEMENT REGULATORY PROTEIN.
GN   CRRY.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=96006570; PubMed=7590969;
RA   Quigg R.J., Lo C.F., Alexander J.J., Sreed A.E., Moxley G. III;
RT   "Molecular characterization of rat Crry: widespread distribution of
RT   two alternative forms of Crry mRNA.";
RL   Immunogenetics 42:362-367(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Dohi N., Sakurada C., Nonaka M., Okada N., Okada H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L36532; AAA91821.1; -.
DR   EMBL; D42115; BAA22548.1; -.
DR   HSSP; P10998; 1VVD.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00084; sushi; 7.
DR   SMART; SM00032; CCP; 7.
SQ   SEQUENCE   559 AA;  61680 MW;  29E10FFA21DB9BCE CRC64;

  Query Match             22.9%;  Score 172.5;  DB 11;  Length 559;
  Best Local Similarity   30.2%;  Pred. No. 3.9e-11;
  Matches   42;  Conservative   21;  Mismatches   55;  Indels   21;  Gaps    5;

Qy       2 ISCDPPPEVKNA------RKPYYSLPIVPGTVLRYTCSPS------YRLIGEKAIFCISE 49
           | |: || : |       |: ::      | |: | |:           : :|| :| | |
Db     159 IPCEIPPSIPNGDFFSPNREDFHY-----GMVVTYQCNTDARGKKLFNLVGEPSIHCTSI 213

Qy      50 NQVHATWDKAPPICESVNKTISCSDPIVPGGFMNKGSKAPFRHGDSVTFTCKANFTMKGS 109
           :     |   || |  :||   |: | |    :   :|: |   | | | |:  | |||
Db     214 DGQVGVWSGPPPQCIELNK---CTPPHVENAVIVSKNKSLFSLRDMVFFRCQDGFMMKGD 270

Qy     110 KTVWCQANEMWGPTALPVC 128
           :|:|::   | |  || |
Db     271 SSVYCRSLNRWEP-QLPSC 288


RESULT 45
Q91YB6
ID   Q91YB6      PRELIMINARY;      PRT:  1230 AA.
AC   Q91YB6;
```

```
DT   02-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   COMPLEMENT INHIBITORY FACTOR H.
GN   CFH.
OS   Rattus norvegicus (Rat)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA   Sandoval F., Goetze O., Schlat G.;
RT   "Rat complement factor H: molecular cloning, sequencing and expression
RT   in tissues and isolated cells.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ420722; CAC67514.1; -.
SQ   SEQUENCE   1236 AA;  140443 MW;  1AC89EFA28232EBF CRC64;

  Query Match             22.9%;  Score 172;  DB 11;  Length 1236;
  Best Local Similarity   37.4%;  Pred. No. 1.1e-10;
  Matches   92;  Conservative   11;  Mismatches   94;  Indels   22;  Gaps    9;

Qy        4 SCDPP      PFVKNARKDYYSLPIVPGIVLRYDCSPSYRLIGEKAIFCISENQVHAIWD 57
            ||| |       |:|  :|  ||      | |: | | | : | |    ) )      |   |
Db      628 SCTQPLDIFIPNCFLKGIKKVIYS     eGDVVEYECKPGFELKGPNKIQC  ---VDGKW1 678

Qy       58 KAPPICESVNKPTSCSGPPVPEMNKGSKAPPPHGDSVTFTCKANPTMKGSKIVWCQA 116
              |||     :|   | | | :  | : | |   |: ||||| |||    ||| |    :|| :
Db      679 IL PLCVEYERK  NGDLPELEHGSV KLSVPPYHHGDSVEFTCPEPFTMIGHAVVFCIS 734

Qy      117 NEMWGPCALPVCPSDFPLE 135
              :    :| : :   ||
Db      735 GRCW  TELPQCVATDQLE 753


Search completed: November 6, 2002, 16:08:26
Job time : 25.5799 secs
```